

Sequence Specific Alkylation of DNA by Polyamide-Chlorambucil Conjugates

Thesis by
Nicholas R. Wurtz

In Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy

California Institute of Technology
Pasadena, California
2002

(Submitted November 27, 2001)

© 2001

Nicholas R. Wurtz

All Rights Reserved

Acknowledgments

I would like to thank my advisor, Professor Peter Dervan, for providing an exciting environment to do research. I would also like to thank the members of my committee, Professors Robert Grubbs, Steve Mayo, and David MacMillan.

While in the Dervan group I have worked with a wonderful group of people. I have immensely enjoyed many conversations with Doan about science, careers and life in general and hope that they continue. Ryan taught me the “Dervan way” and became a good friend. Aileen is perhaps the nicest person that I have ever encountered and is also a patient instructor from whom I learned much. Jason Belitsky has a wonderful dry wit and can find a silver lining in most any lab result and will make a good professor someday. I enjoyed trading investment advice with Clay and admire his determination. Victor provided clever insight that made me smile on many occasions. Anna always provided good advice. I’ve enjoyed getting to know Phillip and his family in the past year. It was a pleasure to have another father in lab to laugh with about the exploits of our sons. It has been an honor to collaborate with Joel Pomerantz and Matt Porteus, who also had the patience to answer my many questions about biology. I have enjoyed sharing a lab with Ryan, Ulf, Eric, Shane and Nick. Thank you for putting up with my many messes.

Most importantly, I would like to thank my family. I would not be here without the support of my parents. I contribute my healthy outlook on life to my wife and son, Heather and Alexander, both of whom I love dearly. They made me smile even after the most frustrating days in lab and helped me keep things in perspective. I am blessed to have such loving wife with whom to share my life.

Abstract

Small molecules that bind selectively to a DNA sequence in the human genome are potentially useful tools for molecular biology and human medicine. Polyamides containing N-methylimidazole (Im) and N-methylpyrrole (Py) are small molecules that bind DNA according to a set of “pairing rules” with affinities and specificities similar to many naturally occurring DNA binding proteins. Two alternate routes of synthesis of polyamides are presented that allowed improved synthesis of commonly used motifs as well as access to polyamides with truncated tails which increase the number of DNA sequences targetable by polyamides with high affinity. Py-Im polyamides may offer a general approach to chemical regulation of gene expression provided inhibition of DNA binding for a variety of transcription factor families can be achieved. Two successful strategies are shown for the inhibition of DNA binding by two such proteins, GCN4 and NF- κ B. An alternate method of gene regulation would be inhibition of transcription during the elongation phase by site-specific alkylation in the coding region of genes. A new class of compounds is created by the attachment of a bis (dichloroethylamino)-benzene nonspecific DNA alkylator to polyamides, which alkylate DNA with high yield and specificity. The biological activity and cell permeability of one of these compounds in several systems is studied.

Table of Contents

	page
Acknowledgments.....	iii
Abstract.....	iv
Table of Contents.....	v
List of Figures and Tables.....	vi
 CHAPTER ONE: Introduction.....	 1
 CHAPTER TWO: Fmoc Solid Phase Synthesis of Polyamide Containing Pyrrole and Imidazole Amino Acids	 16
 CHAPTER THREE: Solid Phase Synthesis of Six-ring Polyamides with Truncated C-Terminal Tails	 32
 CHAPTER FOUR: Inhibition of Major Groove DNA Binding bZIP Proteins by Positive Patch Polyamides.....	 55
 CHAPTER FIVE: Inhibition of DNA Binding by NF- κ B with Pyrrole- Imidazole Hairpin Polyamides	 85
 CHAPTER SIX: Sequence Specific Alkylation of DNA by Hairpin Pyrrole-Imidazole Polyamide Conjugates	 104
 CHAPTER SEVEN: Alkylation, Toxicity and Cell Permeability of Chlorambucil-Polyamide Conjugates.....	 129

List of Figures and Tables

CHAPTER ONE	page
Figure 1.1 Structure of double-helical B-form DNA.....	3
Figure 1.2 Representative X-ray crystal structures of DNA binding proteins.....	4
Figure 1.3 Structures of DNA-binding molecules from natural sources.....	5
Figure 1.4 Schematic representation of the polyamide pairing rules.....	6
Figure 1.5 X-ray crystal structure polyamide homodimer.....	7
Figure 1.6 Representative motifs for polyamide:DNA recognition.....	8
Figure 1.7 Inhibition of gene transcription by hairpin polyamides.....	9
 CHAPTER TWO	
Figure 2.1 Structures of polyamides prepared by solid phase synthesis	19
Figure 2.2 Synthesis of Fmoc-protected pyrrole and imidazole monomers	20
Figure 2.3 New route to Boc-protected imidazole monomer	21
Figure 2.4 Fmoc solid phase synthesis of polyamides	22
Figure 2.5 Polyamide synthesis via reduction of aryl nitro groups	23
 CHAPTER THREE	
Figure 3.1 H-bonding model for six-ring polyamides.....	35
Figure 3.2 Structures of six-ring polyamides with modified tails.....	37
Figure 3.3 Synthesis of polyamide with primary amide tail on Rink resin.....	38
Figure 3.4 Synthesis of polyamides with truncated tails on oxime resin.....	40
Figure 3.5 Dnase I footprinting experiments.....	42
Figure 3.6 Summary of end effects by truncated polyamides.....	46
 Table 3.1 Equilibrium association constants for six-ring hairpins.....	44

CHAPTER FOUR

Figure 4.1	Structures of N-aminoalkylpyrrole polyamides	59
Figure 4.2	Synthesis of Boc-Py(C ₃ OH)-acid monomer	61
Figure 4.3	Synthesis of N-aminoalkylpyrrole polyamides.....	62
Figure 4.4	Dnase I footprinting experiments.....	63
Figure 4.5	N-aminoalkylpyrrole polyamides for GCN4 inhibition	66
Figure 4.6	GCN4 inhibition by N-aminoalkylpyrrole polyamides.....	68
Figure 4.7	Levels of inhibition for selected polyamides.....	69
Figure 4.8	GCN4 inhibition at shifted polyamide target sites	71

Table 4.1	Equilibrium association constants.....	65
-----------	--	----

CHAPTER FIVE

Figure 5.1	X-ray crystal of p50p65 NF- κ B heterodimer	89
Figure 5.2	Polyamides bound to NF- κ B DNA binding site.....	90
Figure 5.3	Synthesis of Boc-Im(C ₃ NHBoc)-acid monomer	91
Figure 5.4	Structures of polyamides targeted to NF- κ B DNA binding site	92
Figure 5.5	DNase I footprinting experiments.....	94
Figure 5.6	Gel mobility shift assays	95
Figure 5.7	Potential explanations for inhibition of NF- κ B binding.....	96
Figure 5.8	Additional polyamides targeted to NF- κ B DNA binding site	97

Table 5.1	Protein targets for polyamides inhibitors.....	88
-----------	--	----

Table 5.2	Equilibrium association constants.....	93
-----------	--	----

CHAPTER SIX

Figure 6.1	H-bond model for polyamide conjugate binding and alkylation.....	107
Figure 6.2	Structures of hairpin polyamide conjugates.....	108
Figure 6.3	pHIV-LTR restriction fragment.....	109
Figure 6.4	Synthesis of polyamide-chlorambucil conjugate	110

Figure 6.5	DNase I footprinting experiment	112
Figure 6.6	Thermal cleavage assay	113
Figure 6.7	Map of alkylation sites and proposed models for alkylation.....	115
Figure 6.8	Inosine substitution for identification of an alkylation adduct	116
Figure 6.9	Alkylation time course	117
Figure 6.10	Alkylation by polyamide homodimer-chlorambucil conjugate	118

CHAPTER SEVEN

Figure 7.1	Chemical structures of studied compounds	132
Figure 7.2	Cross-linking assay	133
Figure 7.3	Alkylation of SV40 DNA and SV40 minichromosomes	134
Figure 7.4	Inhibition of SV40 DNA replication.....	135
Figure 7.5	Comparison of <i>in vitro</i> and <i>in vivo</i> alkylation	137
Figure 7.6	Inhibition of cell growth.....	138
Figure 7.7	Structures of dye-labeled polyamides.....	139
Figure 7.8	Synthesis of dye-chlorambucil polyamide conjugate.....	140
Figure 7.9	Imaging cell trafficking of polyamide-dye conjugates	142
Table 7.1	Cell lines sensitive to chlorambucil-polyamide conjugate.....	138
Table 7.2	Cell lines not sensitive to chlorambucil-polyamide conjugate.....	138

CHAPTER 1

Introduction

Background

The recent sequencing of the human genome has revealed that our genetic information is comprised of 3.2 gigabases of DNA organized into 30,000-40,000 protein-coding genes (1, 2). It is now understood that nature controls the expression of each of these genes by a remarkable three-dimensional switch, that is composed of over 50 proteins assembled on a few hundred specific base pairs of 'promoter' DNA sequence encoded upstream from the RNA polymerase start site and coding region. Each cell contains a copy of the individual's DNA that is essentially identical to that of every other cell, yet each cell has a specialized function that requires it to express certain genes while leaving others silent (3). Misregulation of gene expression resulting in aberrant biochemical behavior in cellular processes is responsible for numerous cancerous and viral disease states (4). Artificial ligands that can sequence specifically bind DNA and restore the natural patterns of gene expression would potentially be valuable tools in human medicine (5).

DNA Recognition

The double helix consists of the two strands of deoxyribose-phosphate polymers that display four heterocycles bases, adenine (A), thymidine (T), guanine (G), and cytosine (C), bound in anti-parallel orientation. In duplex DNA, the two strands are held together via Watson-Crick hydrogen bonding of A, T and G,C base pairs (6). The common B-form of DNA is characterized by a wide (12 Å) and shallow major groove and a narrow (4-6 Å) and deep minor groove (Figure 1.1) (7). In addition, sequence-dependent structural variations, conformational properties, and solvent and counterion

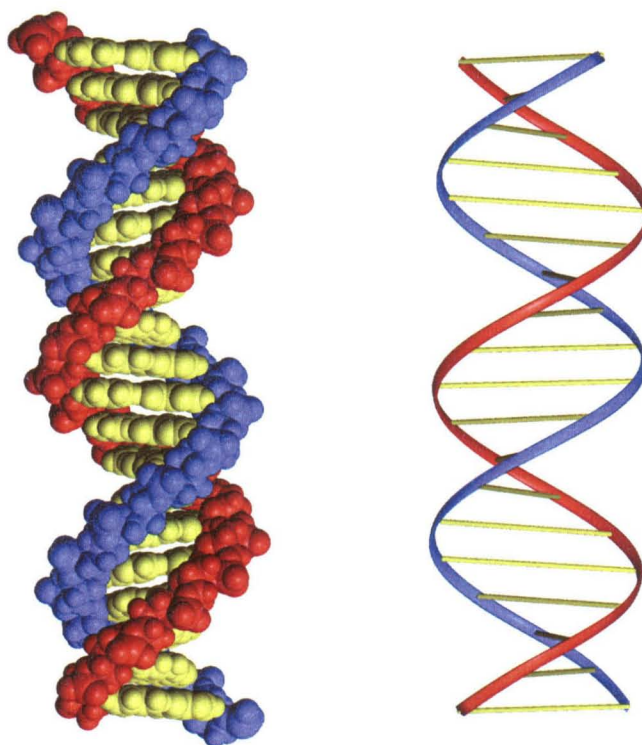


Figure 1.1. CPK model (left) and ribbon representation (right) of B-form duplex DNA. The sugar phosphate backbone is shown in either blue or red. The base pairs are in yellow. (right)

organization can distinguish local DNA structures (7). Individual sequences may be distinguished by the pattern of hydrogen bond donors and acceptors, created by the base pairs.

X-ray crystal and NMR structural analyses of naturally occurring protein-DNA complexes and natural products has provided insight into the design of novel DNA binding molecules (8, 9). Sequence-specificity arises from several methods of DNA interaction: specific hydrogen bonding or van der Waals contacts with functional groups in the grooves, Coulombic attraction to the negatively charged phosphodiester backbone or to the electrostatic potential in the grooves, and/or intercalation of aromatic functional groups between the DNA bases (10). In cells, nature has devised a combinatorial approach for DNA recognition, utilizing a 20 amino acid code to form complex tertiary

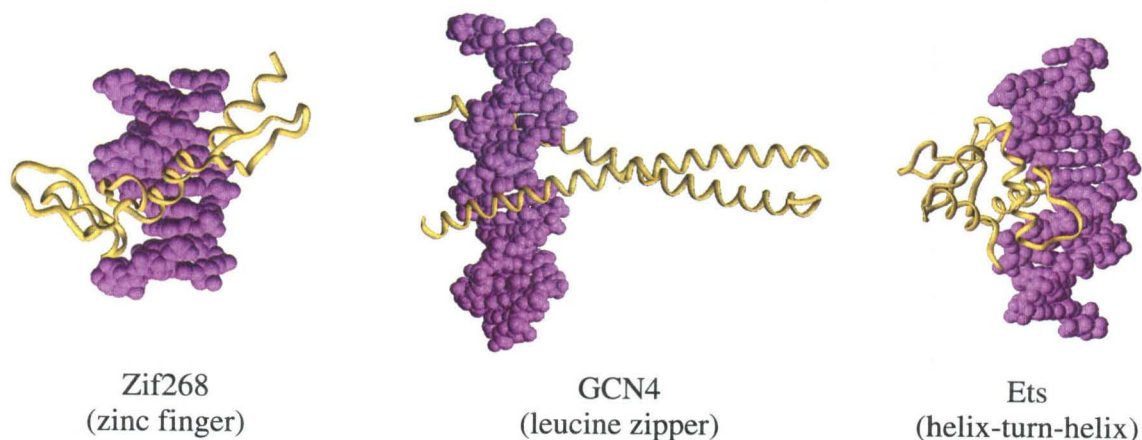
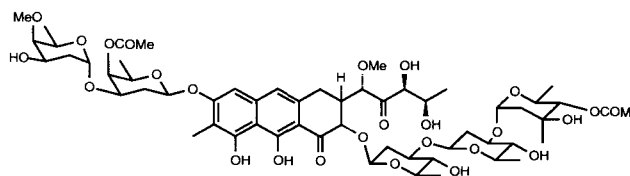


Figure 1.2. Representative X-ray crystal structures from a few families of DNA binding proteins. Proteins contact the DNA via interactions with the bases in the major groove and/or minor groove, as well as the sugar phosphate backbone.

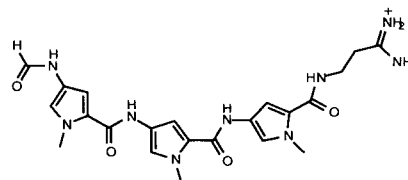
folded structures. DNA binding proteins adopt several structural motifs for sequence-specific recognition including the zinc finger (11), the leucine zipper (12), and the helix-turn-helix motifs (13) (Figure 1.2). Within these complexes, specificity for target sites is achieved through specific noncovalent interactions between the protein side chains and the nucleobases and phosphates of the DNA. However, no single motif exists that represents a general amino acid-base pair code for all DNA sequences (14). Although certain zinc finger-DNA complexes (15, 16) provide a versatile recognition code, the de novo design of proteins for DNA recognition remains challenging due to protein structural diversity and limitations in predicting protein folding.

Small molecules isolated from natural sources are another class of DNA binding ligands that are structurally diverse, as evidenced by consideration of four representative molecules, chromomycin, distamycin, actinomycin D, and calicheamicin (Figure 1.3) (17-20). While these molecules can recognize short sequences of DNA through both intercalation between the bases and groove binding, there is no general recognition code

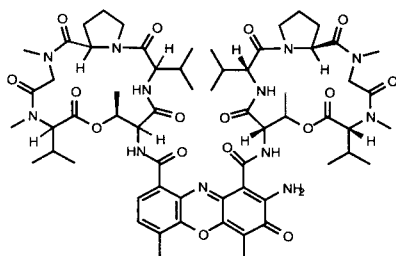
for the readout of specific sequences of DNA. Analogs of calicheamicin have been shown to bind DNA and interfere with transcription factor function; however, oligosaccharides



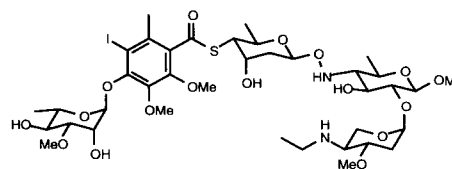
Chromomycin • 5'-GGCT-3'



Distamycin • 5'-(A,T)₅-3'



Actinomycin D • 5'-GC-3'



Calicheamicin oligosaccharide • 5'-TCCT-3'

Figure 1.3. The structures of four small molecules isolated from natural sources.

are not yet amenable to recognition of a broad range of predetermined DNA sequences (21, 22). Similarly, oligodeoxynucleotides which recognize the major groove of double helical DNA via triple helix formation bind a broad sequence repertoire with high affinity and specificity (23, 24). Although oligonucleotides and their analogs have been shown to interfere with gene expression, the triple helix approach is limited to purine tracks and suffers from poor cellular uptake (25, 26).

DNA Recognition by Polyamides.

Over the past two decades, the research efforts led by Prof. Peter B. Dervan of Caltech have resulted in a set of pairing rules to guide polyamide design for sequence-specific recognition of DNA. A binary code has been developed to correlate DNA sequence with side-by-side pairings between pyrrole (Py), imidazole (Im), and 3-

hydroxypyrrole (Hp) carboxamides in the DNA minor groove (Figure 1.4). A pairing of Im opposite Py (Im/Py) targets a G•C base pair, while Py/Im targets C•G. A Py/Py pairing is degenerate, targeting both A•T and T•A base pairs. An Hp opposite a Py (Hp/Py) discriminates T•A from A•T, while Py/Hp targets A•T in preference to T•A and both of these from G•C and C•G. Footprinting, NMR, and X-ray structure studies validate these pairing rules for DNA minor groove recognition (27, 28).

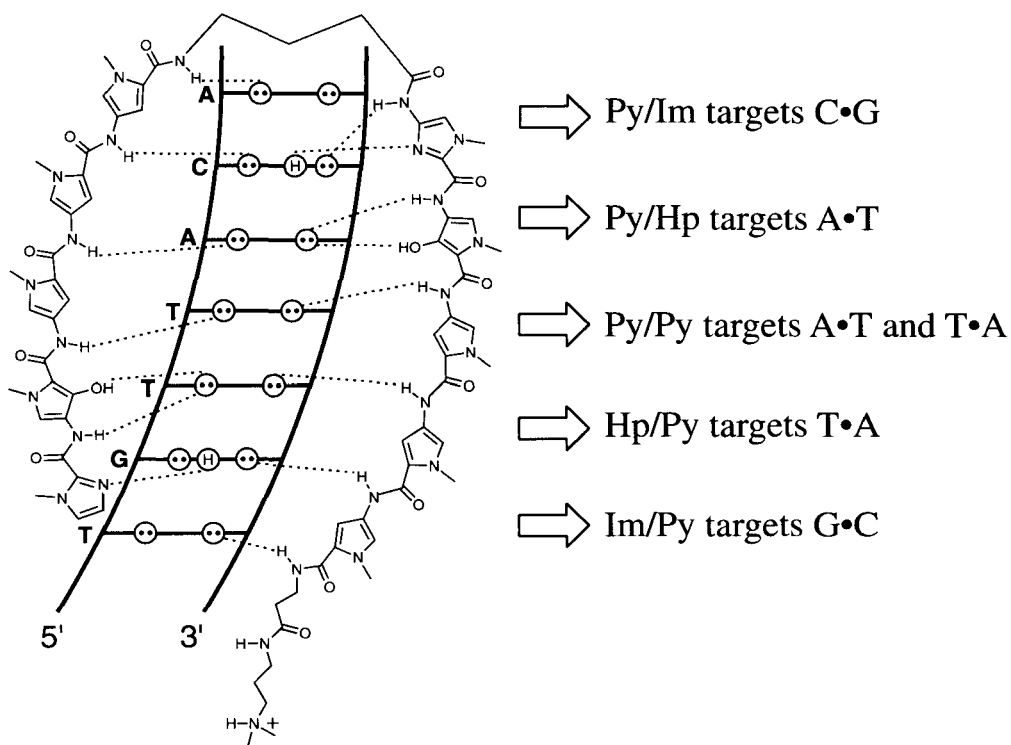


Figure 1.4. A schematic representation of the polyamide pairing rules.

The DNA-binding affinity and specificity of pyrrole-imidazole polyamides have been increased by covalently linking polyamide subunits. Polyamide dimers linked with a γ -aminobutyric acid (γ) turn residue provide a hairpin motif that binds target sites with two to three orders of magnitude enhanced affinity and has the important feature that ring pairings are unambiguously set in place as relative to homodimers which can afford

slipped motifs (29). Increasing the hairpin polyamide subunit lengths to four and five aromatic ring residues gives eight- and ten-ring hairpin polyamides with subnanomolar binding affinities, similar to naturally occurring DNA-binding proteins (30, 31). Studies of polyamide site size limitations suggest that beyond five consecutive rings, the ligand curvature fails to match the pitch of the DNA helix, disrupting the hydrogen bonds and van der Waals interactions responsible for specific polyamide-DNA complex formation (32). Addition of pairings with β -alanine (β) to form β/β , β/Py , and β/Im pairs has extended the targetable binding site size recognizable by the hairpin motif at high affinity (33, 34).

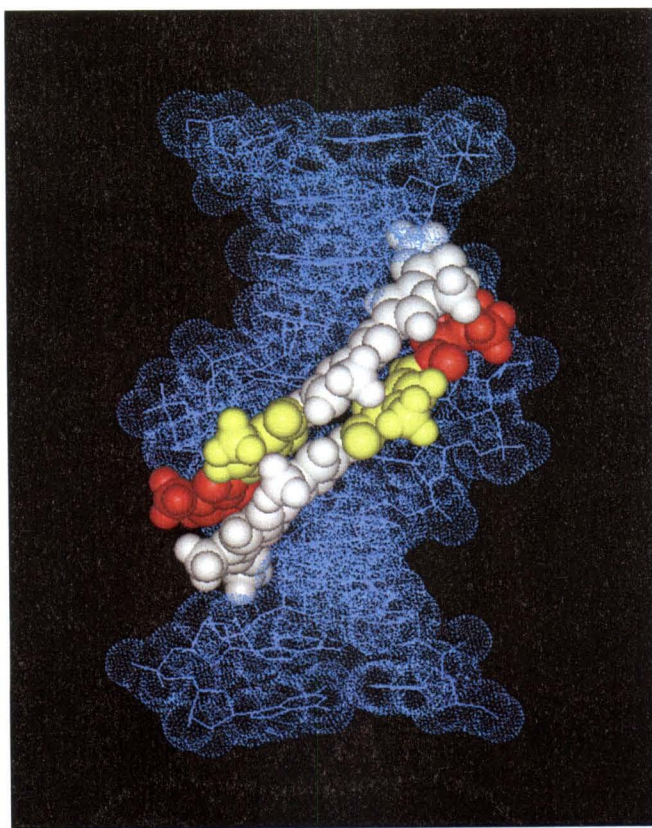


Figure 1.5. Space filling representation of the polyamide dimer ImHpPyPy- β -Dp bound in the minor groove of DNA. The figure was prepared using Insight II software and is derived from a high-resolution X-ray co-crystal structure of a polyamide dimer bound to DNA, which was obtained in collaboration with the Rees group at the California Institute of Technology.

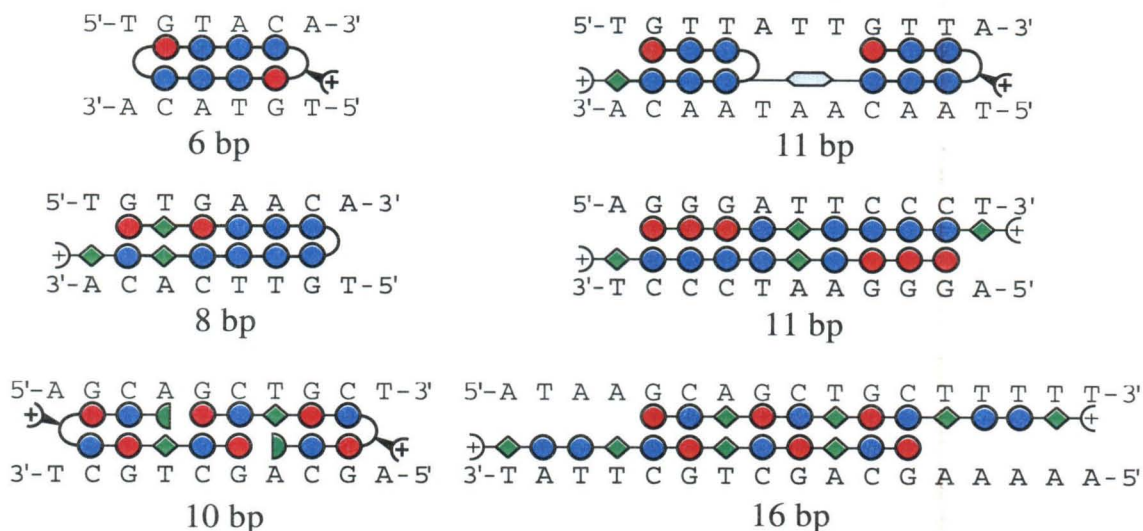


Figure 1.6. Representative motifs for polyamide:DNA recognition. Red and blue circles represent imidazole and pyrrole residues, respectively. β -Alanine and the γ -turn are depicted as green diamonds and curved lines, respectively. The plus sign represents the dimethylaminopropylamide tail.

A variety of motifs for sequence specific recognition using the polyamide scaffold have been successfully developed for target sites from 5-16 bp (Figure 1.6) (29-31, 35-41). For longer sequences, the introduction of a flexible β -alanine allows the polyamide that would otherwise be overwrought relative to the DNA helix to track the minor groove (29, 32, 34, 38, 42). Cooperative 2:1 dimer motifs are particularly attractive for future applications due to the low molecular weight of the monomer units (36). Binding of sites up to 16 bp, potentially sufficient for recognition of a unique sequence in the human genome, have been successfully targeted (37). Closing the ends of a hairpin polyamide with a second γ affords the cycle motif which can increase polyamide affinity (40, 41). The incorporation of a stereospecific α -amino group into the γ turn also increases polyamide affinity and specificity, while allowing for functional moieties to be linked or the construction of a tandem motif by linking the C-terminus of another polyamide to this position (39, 43).

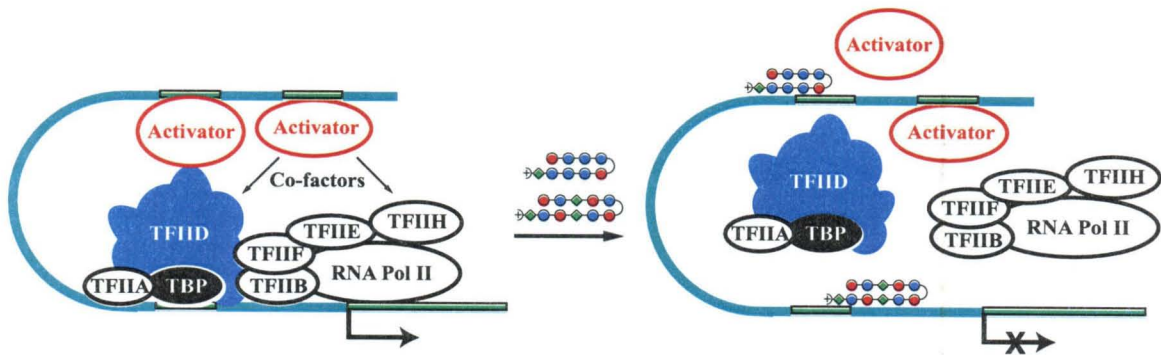


Figure 1.7. Inhibition of gene transcription by hairpin polyamides. A pair of hairpin polyamides targeted to the DNA sequences adjacent to the binding sites for LEF-1 and TBP inhibit assembly of the transcriptional machinery and transcription of the HIV-1 gene in human cell culture.

Gene Regulation by Polyamides

As an initial step toward asking whether cell-permeable small molecule transcription agonists might regulate gene expression in complex organism, regulation of transcription in specific genes by polyamides has been studied. The DNA-binding activity of the 5S RNA specific transcription factor TFIIIA was inhibited by an eight-ring hairpin polyamide that bound within the recognition site of the nine-zinc-finger protein in the minor groove. Transcription of 5S RNA genes by RNA polymerase III was suppressed *in vitro* and in cultured *Xenopus* kidney cells (5). Polyamides have also been shown to downregulate genes transcribed by RNA polymerase II. Two hairpin polyamides were targeted to DNA sequences adjacent to DNA sequences immediately adjacent to the binding site of transcription factors, Ets-1, LEF-1 and TBP in the HIV-1 enhancer/promoter. These synthetic ligands inhibit DNA-binding of each transcription factor and HIV-1 transcription in cell-free assays. Treatment of isolated human peripheral blood lymphocytes with the two hairpin polyamides demonstrated inhibition of viral replication (33). In an effort to create an artificial activator, capable of upregulating gene

expression, a hairpin polyamide was tethered to a short peptide (20mer) activation domain (AH). The conjugate successfully upregulated transcription in a cell-free system (44).

Scope of this Work

This thesis describes work including development of new methodologies for synthesizing polyamides, application of new polyamide motifs toward protein inhibition and synthesis and evaluation of sequence specific alkylating agents.

Chapters 2 and 3 describe two alternative synthesis routes to access polyamides that complement previously reported Boc solid phase synthesis methodology. In Chapter 2 details a Fmoc solid phase synthesis method that is amenable to both manual and machine-assisted strategies. Purities and yields were equal or better to those reported with Boc-chemistry. Chapter 3 introduces access of polyamides with truncated tails by synthesis on Rink resin using Fmoc monomers or using Boc monomers on oxime resin. The new tails allow binding of all four base pairs at the tail position expanding the range of DNA sequences that can be targeted with high affinity by hairpin polyamides.

Chapters 4 and 5 report the inhibition of protein:DNA complex formation by hairpin polyamides. Chapter 4 presents the inhibition of major groove-binding GCN4 by polyamides by targeting specific protein-DNA contacts along the phosphate backbone with positively charged alkylamino substituents off the pyrrole ring. Relocation of the positive charge from the ring to the tail resulted in a tenfold increase in binding affinity with uncompromised specificity. Chapter 5 demonstrates the inhibition of DNA-binding by NF- κ B. Inhibition of the p50p65 NF κ B heterodimer is successful only when a

polyamide is targeted to the 5'-GGGAC-3' subsite. A new polyamide motif is reported that binds a 3'-GGGG-5' site with subnanomolar affinity.

Chapters 6 and 7 describe the creation of a new class of sequence specific alkylators by the combination of DNA binding polyamides with the classic, nonspecific alkylator, chlorambucil. Chapter 6 describes the synthesis of a polyamide-chlorambucil conjugate targeted to the HIV-1 promoter region. The yield and specificity of alkylation is determined by thermal cleavage assays on a 241 bp restriction fragment. Chapter 7 describes pharmacological data of polyamide-chlorambucil conjugates. Alkylation of the HIV-1 promoter in genomic DNA in cells and compared to alkylation of naked genomic DNA. The varying toxicity of the compounds to various cell lines and effects of cell cycles is reported.

References

1. Consortium, I. H. G. S. (2001) *Nature* 409, 860.
2. Venter, J. C. et. al. (2001) *Science* 291, 1304.
3. Roeder, R. G. (1996) *Trends in Biochemical Sciences* 9, 327.
4. Tjian, R. (1995) *Scientific American* 2, 54.
5. Gottesfeld, J. M., Neely, L., Trauger, J. W., Baird, E. E., and Dervan, P. B. (1997) *Nature* 387, 202-205.
6. Dickerson, R. E., Drew, H. R., Conner, B. N., Wing, M., Fratini, A. V., and Kopka, M. L. (1982) *Science* 216, 475.
7. Saenger, W. (1984) *Principles of Nucleic Acid Structure*, Springer-Verlag, New York.
8. Zimmer, C., and Wahnaert, U. (1986) 47, 31-112.
9. Steitz, T. A. (1990) *Quarterly Review of Biophysics* 23, 205.
10. Pabo, C. O., and Sauer, R. T. (1992) *Annual Review of Biochemistry* 61, 1053-1095.
11. Pavletich, N. P., and Pabo, C. O. (1991) *Science* 252, 809.
12. Ellenberger, T. E., Brandl, C. J., Struhl, K., and Harrison, S. C. (1992) *Cell* 71, 1223.
13. Feng, J. A., Johnson, R. C., and Dickerson, R. E. (1994) *Science*, 348.
14. Kissinger, C. R., Liu, B., Martin-Blanco, E., Kornberg, T. B., and Pabo, C. O. (1990) *Cell* 63, 579.
15. Greisman, H. A., and Pabo, C. O. (1997) *Science* 275, 657-661.

16. Segal, D. J., and Barbas, C. F. (2000) *Current Opinion in Chemical Biology* 4, 34-39.
17. Coll, M., Frederick, C. A., Wang, A. H. J., and Rich, A. (1987) *Proceedings of the National Academy of Sciences of the United States of America* 84, 8385-8389.
18. Gao, X. L., Mirau, P., and Patel, D. J. (1992) *Journal of Molecular Biology* 223, 259-279.
19. Kamitori, S., and Takusagawa, F. (1992) *Journal of Molecular Biology* 225, 445-456.
20. Paloma, L. G., Smith, J. A., Chazin, W. J., and Nicolaou, K. C. (1994) *Journal of the American Chemical Society* 116, 3697-3708.
21. Ho, S. N., Boyer, S. H., Schreiber, S. L., Danishefsky, S. J., and Crabtree, G. R. (1994) *Proceedings of the National Academy of Sciences of the United States of America* 91, 9203-9207.
22. Liu, C., Smith, B. M., Ajito, K., Komatsu, H., GomezPaloma, L., Li, T. H., Theodorakis, E. A., Nicolaou, K. C., and Vogt, P. K. (1996) *Proceedings of the National Academy of Sciences of the United States of America* 93, 940-944.
23. Moser, H. E., and Dervan, P. B. (1987) *Science* 238, 645-650.
24. Thuong, N. T., and Helene, C. (1993) *Angewandte Chemie-International Edition in English* 32, 666-690.
25. Maher, L. J., Dervan, P. B., and Wold, B. (1992) *Biochemistry* 31, 70-81.
26. Nielsen, P. E. (1997) *Chemistry-a European Journal* 3, 505-508.
27. Dervan, P. B. (2001) *Bioorganic & Medicinal Chemistry* 9, 2215-2235.

28. Dervan, P. B., and Bürli, R. W. (1999) *Current Opinion in Chemical Biology* 3, 688-693.
29. Trauger, J. W., Baird, E. E., Mrksich, M., and Dervan, P. B. (1996) *Journal of the American Chemical Society* 118, 6160-6166.
30. Trauger, J. W., Baird, E. E., and Dervan, P. B. (1996) *Nature* 382, 559-561.
31. Turner, J. M., Baird, E. E., and Dervan, P. B. (1997) *Journal of the American Chemical Society* 119, 7636-7644.
32. Kielkopf, C. L., Baird, E. E., Dervan, P. D., and Rees, D. C. (1998) *Nature Structural Biology* 5, 104-109.
33. Dickinson, L. A., Gulizia, R. J., Trauger, J. W., Baird, E. E., Mosier, D. E., Gottesfeld, J. M., and Dervan, P. B. (1998) *Proceedings of the National Academy of Sciences of the United States of America* 95, 12890-12895.
34. Turner, J. M., Swalley, S. E., Baird, E. E., and Dervan, P. B. (1998) *Journal of the American Chemical Society* 120, 6219-6226.
35. Trauger, J. W., Baird, E. E., and Dervan, P. B. (1996) *Chemistry & Biology* 3, 369-377.
36. Trauger, J. W., Baird, E. E., and Dervan, P. B. (1998) *Angewandte Chemie-International Edition* 37, 1421-1423.
37. Trauger, J. W., Baird, E. E., and Dervan, P. B. (1998) *Journal of the American Chemical Society* 120, 3534-3535.
38. Swalley, S. E., Baird, E. E., and Dervan, P. B. (1997) *Chemistry-a European Journal* 3, 1600-1607.

39. Herman, D. M., Baird, E. E., and Dervan, P. B. (1999) *Chemistry-a European Journal* 5, 975-983.
40. Cho, J., Parks, M. E., and Dervan, P. B. (1995) *Proceedings of the National Academy of Sciences, USA* 92, 10389-10392.
41. Herman, D. M., Turner, J. M., Baird, E. E., and Dervan, P. B. (1999) *Journal of the American Chemical Society* 121, 1121-1129.
42. deClairac, R. P. L., Seel, C. J., Geierstanger, B. H., Mrksich, M., Baird, E. E., Dervan, P. B., and Wemmer, D. E. (1999) *Journal of the American Chemical Society* 121, 2956-2964.
43. Herman, D. M., Baird, E. E., and Dervan, P. B. (1998) *Journal of the American Chemical Society* 120, 1382-1391.
44. Mapp, A. K., Ansari, A. Z., Ptashne, M., and Dervan, P. B. (2000) *Proceedings of the National Academy of Sciences of the United States of America* 97, 3930.

CHAPTER 2

Fmoc Solid Phase Synthesis of Polyamides Containing Pyrrole and Imidazole Amino Acids

The text of this chapter was taken in part from a publication co-authored with James M. Turner, Eldon E. Baird & Prof. Peter B. Dervan.

(N. R. Wurtz, J. M. Turner, E. E. Baird & P. B. Dervan *Org. Lett.*, **2001**, 3, 1201-1203.)

Abstract

Polyamides containing N-methylimidazole (Im) and N-methylpyrrole (Py) amino acids are synthetic ligands that have an affinity and specificity for DNA comparable to many naturally occurring DNA-binding proteins. A machine assisted Fmoc solid phase synthesis of polyamides has been optimized to afford high stepwise coupling yields (>99%). Two monomer building blocks, Fmoc-Py acid and Fmoc-Im acid, were prepared in multigram scale. Cleavage by aminolysis followed by HPLC purification affords up to 200 mg quantities of polyamide with purities and yields greater than or equal to those reported using Boc chemistry. Additionally, an improved synthesis of Boc-Im acid is reported. A broader set of reaction conditions will increase the number and complexity of minor groove binding polyamides that may be prepared, and help ensure compatibility with many commercially available peptide synthesizers.

Introduction

Small molecules specifically targeted to any predetermined DNA sequence would be useful tools in molecular biology and potentially in human medicine. Polyamides containing *N*-methylpyrrole (Py), 3-hydroxypyrrole (Hp) and *N*-methylimidazole (Im) are synthetic ligands that have an affinity and specificity for DNA comparable to naturally occurring DNA binding proteins. DNA recognition depends on side-by-side aromatic amino acid pairings oriented N→C with respect to the 5'→3' direction of the DNA helix in the minor groove. An antiparallel pairing of Im opposite Py (Im/Py pair) recognizes a G•C base pair, while a Py/Im combination recognizes C•G. A Hp/Py pair prefers T•A over A•T. A Py/Py pair is degenerate and recognizes either an A•T or T•A base pair (*1*).

The large repertoire of pyrrole-imidazole polyamides synthesized hallmarks the development of Boc-chemistry solid phase polyamide synthesis (*2*). However, to ensure compatibility with most commercially available peptide synthesizers, we set out to develop an alternative solid phase synthesis incorporating a different protecting group that would provide a route to polyamides with comparable purity and recovery to Boc chemistry. The synthetic schemes made available by this new chemistry could make available a greater variety of bifunctional polyamide conjugates and derivatives. As a minimum first step, we report here monomer synthesis and general protocols for machine-assisted Fmoc-chemistry solid phase synthesis of pyrrole-imidazole polyamides, ImPyPyPy- γ -PyPyPyPy- β -Dp **1** and ImPyPyPy- γ -ImPyPyPy- β -Dp **2** (Figure 2.1), which are identical or superior in recovery, purity, and coupling times to Boc-chemistry protocols.

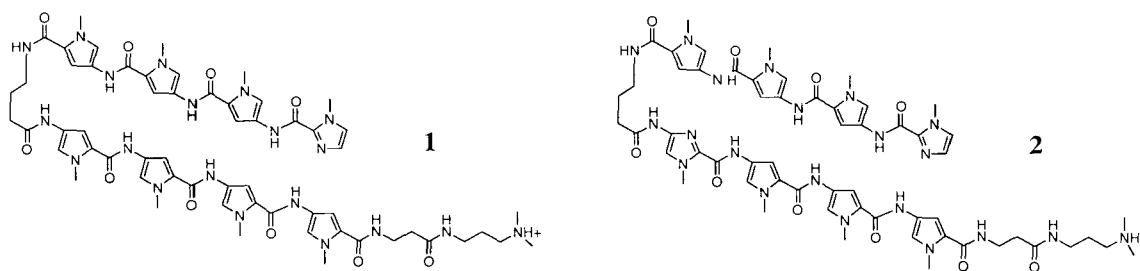


Figure 2.1. ImPyPyPy- γ -PyPyPyPy- β -Dp (top) and ImPyPyPy- γ -PyPyPyPy- β -Dp (bottom) prepared by multistep solid phase synthesis.

The synthesis of Fmoc-Py acid has previously been reported (3, 4). We report a convenient multigram scalable synthesis of Fmoc-Py acid with a higher overall yield without the use of column chromatography, as well as the first reported synthesis of Fmoc-Im acid (Figure 2.2).

Results and Discussion

Pyrrole monomer precursors **3a** and **4a** were synthesized using literature protocols (2). Previously reported 2-(trichloroacetyl)-1-methylimidazole **3b** was converted to *tert*-butyl 1-methylimidazole-2-carboxylate **4b** using modifications of previously published procedures (5). Compound **3b** in acetic anhydride was cooled to 0° C followed by the addition of fuming nitric acid and sulfuric acid over 2 h. Crystallization provided 528 g of **4b** in 56% yield. *tert*-Butyl esters **5a** and **5b** were synthesized from **4a** and **4b** with *tert*-butoxide in refluxing *tert*-butanol yielding **5a** and **5b** in 95% and 84% respectively. Reduction of **5a** and **5b** in DMF in the presence of 10% Pd/C under 500 psi of hydrogen proceeds to completion in three hours. After filtering the reaction mixture through a pad of Celite, Fmoc-Cl and diisopropylethylamine were added to provide **6a** and **6b** in 66% and 35% yield, respectively. The *tert*-butyl esters were deprotected with TiCl₄ in

dichloromethane at 0° C to prevent the decarboxylation, which was observed with standard trifluoroacetic acid deprotection. Precipitation with 1.0 M HCl afforded the final products **7a** and **7b** in 88% and 86% yield, respectively. The overall yield from commercially available starting materials for the Fmoc-Py acid and Fmoc-Im acid are 33% and 11%, respectively.

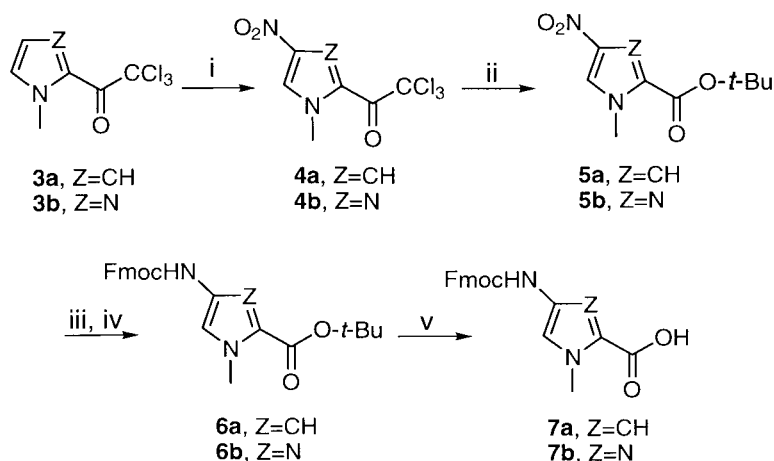


Figure 2.2. (i) 90% nitric acid, Ac₂O; (ii) NaOMe/MeOH; (iii) NaOt-Bu/t-BuOH, (iv) 500 psi of H₂, 10% Pd/C, DMF; (v) Fmoc-Cl, DIEA; (vi) TiCl₄, CH₂Cl₂.

Additionally, the first two steps of the Fmoc-Im acid synthesis provide a safer, higher yielding route to the Boc-Im acid monomer. The original published synthesis of the Boc-Im acid required acylation of N-methyl imidazole with ethyl chloroformate, followed by a highly exothermic nitration at the 4- and 5- positions in sulfuric acid and fuming nitric acid. The isomers were separated by a difficult recrystallization to provide the desired ethyl 1-methyl-4-nitroimidazole-2-carboxylate **8**. In the new synthetic route, nitration of trichloroacetone **3b** in acetic anhydride provides the desired isomer, 4-nitro-2-(trichloroacetyl)-1-methyl-imidazole **4b**, by milder nitration conditions. The product was esterified with sodium ethoxide in ethanol to yield ethyl 1-methyl-4-nitroimidazole-2-carboxylate **8** in 95% yield (Figure 2.3). This intermediate was converted in three steps

into the desired Boc-Im acid **9** using previously reported reactions(2) in an improved overall yield of 18%, compared to the 8% overall yield in the original synthesis.

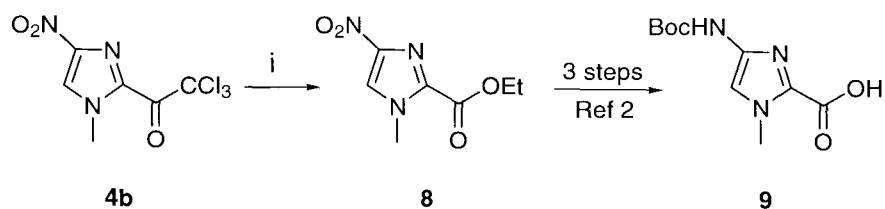


Figure 2.3. (i) NaOMe/MeOH

For Fmoc solid phase synthesis, the polyamide was attached to an insoluble matrix by a linkage, which is cleaved in a single step, introducing a positive charge into the polyamide. Fmoc- β -alanine-Wang resin is commercially available in appropriate substitution levels (0.86 mmol/g). Manual solid phase synthesis uses HBTU/DIEA activation protocols as previously reported (2).

Fmoc-chemistry machine-assisted solid phase polyamide synthesis protocols were modified from the Boc-chemistry protocols reported for use on an ABI 430A peptide synthesizer. Coupling cycles for machine-assisted synthesis are rapid, 180 min per residue, and require no special precautions beyond those used for ordinary solid phase peptide synthesis. Machine-assisted solid phase synthesis of a pyrrole-imidazole polyamide consists of a DCM wash, an NMP wash, removal of the Fmoc group with piperidine/NMP, an NMP wash, a DCM wash, an NMP wash, addition of activated monomer, coupling for 180 minutes, capping with acetic anhydride, taking a resin sample for analysis, and a final NMP wash.

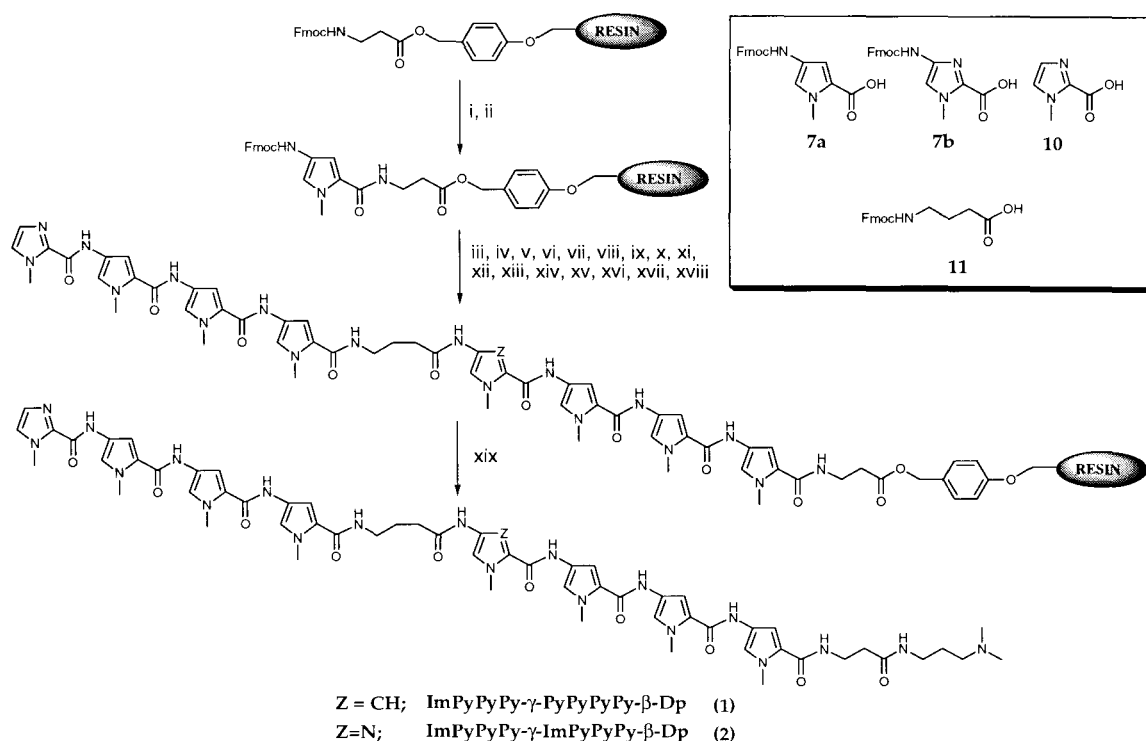


Figure 2.4. Solid phase synthetic scheme for polyamides **1** and **2** starting from commercially available Fmoc-β-alanine-Wang resin. (i) 20% piperidine/NMP; (ii) Fmoc-Py acid, HBTU, DIEA; (iii) 20% piperidine/NMP; (iv) Fmoc-Py acid, HBTU, DIEA; (v) 20% piperidine/NMP; (vi) Fmoc-Py acid, HBTU, DIEA; (vii) 20% piperidine/NMP; (viii) Fmoc-Py acid (for **1**) or Fmoc-Im acid (for **2**), HBTU, DIEA; (ix) 20% piperidine/NMP; (x) Fmoc-γ-aminobutyric acid, HBTU, DIEA; (xi) 20% piperidine/NMP; (xii) Fmoc-Py acid, HBTU, DIEA; (xiii) 20% piperidine/NMP; (xiv) Fmoc-Py acid, HBTU, DIEA; (xv) 20% piperidine/NMP; (xvi) Fmoc-Py acid, HBTU, DIEA; (xvii) 20% piperidine/NMP; (xviii) Im-acid, HBTU, DIEA; (xix) *N,N*-dimethylaminopropylamine, 55° C.

Because the aromatic amine of the pyrrole and imidazole do not react in the quantitative ninhydrin test, stepwise cleavage of a sample of resin and analysis by HPLC is used to indicate that high stepwise yields (>99%) are routinely achieved. We note that coupling of the imidazole amine with Fmoc-Py-OBt ester was not satisfactory. However, acylation with Boc-Py anhydride ester (synthesized *in situ* using DCC, DMAP in DCM) proceeds to completion within 3 h (2).

ImPyPyPy-γ-PyPyPyPy-β-Dp **1** and ImPyPyPy-γ-ImPyPyPy-β-Dp **2** were prepared in 19 steps using the protocols described in the experimental section (Figure 2.4). The yield of each individual coupling step was established as >99% by HPLC

analysis. The resin was cleaved in high yield (>90%) by aminolysis with (*N,N*,*N*,*N*-dimethylamino)propylamine. A single HPLC purification of the eight-ring polyamide was sufficient to obtain a recovery of 38% and 9% for **1** and **2**, respectively, and a final purity for both compounds greater than 98% as determined by a combination of analytical HPLC, ^1H NMR, and mass spectroscopy.

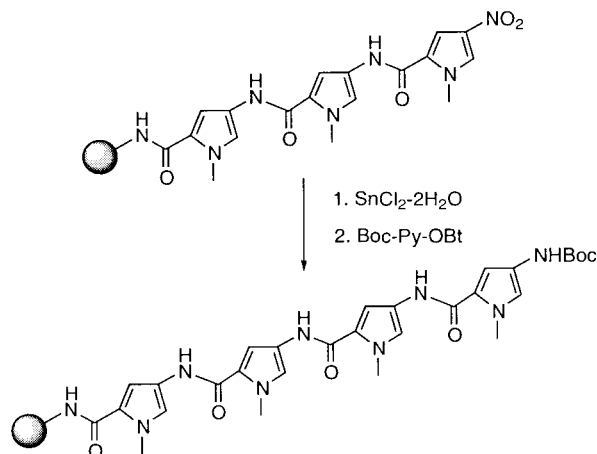


Figure 2.5. Synthesis of *N*-methyl pyrrole polyamides by reduction of aryl-nitro groups on solid phase by tin (II) chloride dihydrate.

Since other synthetic routes may be useful for accessing highly functionalized polyamides, synthesis of polyamides by reduction of aryl nitro groups was investigated. Aryl nitro groups can be selectively reduced to amines on solid phase by tin (II) chloride dihydrate (6-9). This route is advantageous since 4-nitropyrrole carboxylates are accessible at high yields from hydrolysis of the ethyl esters and provide a highly stable amine synthon at the 4-position. In this case, the desired 4-nitropyrrole carboxylate was observed to be a side product of the synthesis of **5a** and isolated by acidifying and extracting the aqueous washes. The monomer coupled readily to a deprotected pyrrole on resin with HBTU/HOBt (Figure 4.5). Reduction of the nitro group with 1.5 M tin (II) chloride dihydrate in dimethylformamide/acetonitrile (7:1) proceeded to completion in

five days at 37° C with little decomposition observed. Deprotection was achieved in 12 hours at 60° C with 2.0 M tin (II) chloride with some decomposition observed. This route of synthesis was not pursued in greater detail since the reduction was slow when compared to Boc or Fmoc deprotection and offers no advantages over either of these routes for current synthetic applications.

Conclusion

Pyrrole-imidazole polyamides are a promising class of compounds used to target predetermined DNA sequences. The synthetic schemes made available by Fmoc-chemistry solid phase will allow a greater variety of bifunctional conjugates and derivatives to be synthesized.

Experimental

Materials. Fmoc- β -alanine-p-benzyloxybenzyl alcohol resin (Fmoc- β -Wang-resin), 2-(1*H*-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU), and Boc- γ -Aminobutyric Acid were purchased from Peptides International. N,N-Diisopropylethylamine (DIEA), piperidine, and N-methylpyrrolidone (NMP) were purchased from Applied Biosystems. Dichloromethane (DCM) and triethylamine (TEA) were reagent grade from EM, 9-Fluorenylmethyl chloroformate (Fmoc-Cl), thiophenol (PhSH), (dimethylamino)propylamine, trichloroacetylchloride, *N*-methylpyrrole, and *N*-methylimidazole were from Aldrich, and trifluoroacetic acid (TFA) was from Halocarbon. All reagents were used without further purification.

^1H NMR spectra were recorded on a General Electric-QE NMR spectrometer at 300 MHz in $\text{DMSO-}d_6$, with chemical shifts reported in parts per million relative to residual solvent. UV spectra were measured in water on a Hewlett-Packard Model 8452A diode array spectrophotometer. Matrix-assisted, laser desorption/ionization time of flight mass spectrometry (MALDI-TOF) electrospray mass spectrometry were performed at the Protein and Peptide Microanalytical Facility at the California Institute of Technology. HPLC analysis was performed on either a HP 1090M analytical HPLC or a Beckman Gold system using a RAINEN C_{18} , Microsorb MV, $5\mu\text{m}$, 300×4.6 mm reversed phase column in 0.1% (wt/v) TFA with acetonitrile as eluent and a flow rate of 1.0 mL/min, gradient elution 1.25% acetonitrile/min. Preparatory reversed phase HPLC was performed on a Beckman HPLC with a Waters DeltaPak 25×100 mm, $100\mu\text{m}$ C18 column equipped with a guard, 0.1% (wt/v) TFA, 8.0 mL/min, 0.25% acetonitrile/min. 18MW water was obtained from a Millipore MilliQ water purification system, and all buffers were $0.2\mu\text{m}$ filtered.

Monomer Synthesis.

***tert*-Butyl 4-Nitro-1-methylpyrrole-2-carboxylate (5a).** To a suspension of 4-nitro-2-(trichloroacetyl)-1-methylpyrrole (**4a**) (50 g, 19.5 mmol) in 500 ml of *tert*-butyl alcohol was added sodium *tert*-butoxide (25 g) over 1 hour. The suspension was refluxed under argon for 5 hours. The mixture was quenched with 500 ml of water and extracted with chloroform (3×500 ml). The chloroform was concentrated *in vacuo* to provide **5a** (39.4 g, 94.6% yield). TLC (3:1 hexanes: ethyl acetate) R_f 0.52 ^1H NMR ($\text{DMSO-}d_6$) δ 8.18 (d, 1 H, $J = 2.4$), 7.18 (d, 1 H, $J = 2.4$), 3.97 (s, 3 H), 1.47 (s, 9 H). FABMS m/e 226.096 (226.095 calcd for $\text{C}_{10}\text{H}_{14}\text{N}_2\text{O}_4$).

***tert*-Butyl 4-[(9-Fluorenylmethoxycarbonyl)amino]-1-methylpyrrole-2-carboxylate (6a).** To a solution of *tert*-Butyl 4-Nitro-1-methylpyrrole-2-carboxylate (**5a**) (20.0 g, 88.4 mmol) in 100 ml of DMF was added 10% Pd/C (2 g) in 20 ml of DMF. The mixture was vigorously stirred for 2 h under 500 psi of hydrogen, filtered through Celite, and the Celite was washed with DMF (1 × 200 ml). The two filtrates were combined, Fmoc-Cl (25.3 g, 92.8 mmol) and DIEA (35 ml, 200 mmol) were added and the solution was stirred for 10 h. Water (250 ml) was added, and the mixture extracted with ethyl ether (2 × 500 ml). The ether was dried (sodium sulfate) and concentrated *in vacuo* to ~200 ml. Hexanes was added to precipitate a brown solid which was washed with cold methanol/H₂O to yield **6a** as a white solid (24.3 g, 66% yield). TLC (3:1 hexanes/ethyl acetate) *R_f* 0.39 ¹H NMR (DMSO-*d*₆) δ 9.45 (s, 1 H), 7.92 (m, 2 H), 7.71 (m, 2 H), 7.44 (m, 2 H), 7.36 (m, 2 H), 7.03 (d, 1 H, *J* = 2.4), 6.63 (d, 1 H, *J* = 2.4), 4.45 (d, 2 H, *J* = 6.3), 4.31 (t, 1 H, *J* = 6.3), 3.78 (s, 3 H), 1.50 (s, 9 H). FABMS *m/e* 418.189 (418.189 calcd for C₂₅H₂₆N₂O₄).

4-[(9-Fluorenylmethoxycarbonyl)amino]-1-methylpyrrole-2-carboxylic acid (7a). *tert*-Butyl 4-[(9-Fluorenylmethoxycarbonyl)amino]-1-methylpyrrole-2-carboxylate (**6a**) (5.0 g, 12.0 mmol) was dissolved in 100 ml of dichloromethane and cooled to 0° C. TiCl₄ (25 ml, 1.0 M in dichloromethane) was added dropwise to the solution. The mixture was stirred for 30 min. 1 M HCl (250 ml) was cooled to 4 °C and added dropwise. The resulting white precipitate (**7a**) was collected by vacuum filtration and washed with a small amount of cold water and dried *in vacuo* (3.8 g, 88% yield). TLC (1:1 hexanes/ethyl acetate) *R_f* 0.15 ¹H NMR (DMSO-*d*₆) δ 12.16 (bs, 1 H), 9.45 (s, 1 H), 7.92 (m, 2 H), 7.74 (m, 2 H), 7.44 (m, 2 H), 7.36 (m, 2 H), 7.07 (d, 1 H, *J* = 2.4), 6.64 (d, 1 H, *J* = 2.4), 4.46 (d, 2 H, *J* = 6.3), 4.30 (t, 1 H, *J* = 6.3), 3.80 (s, 3 H). FABMS *m/e* 362.127 (362.126 calcd for C₂₁H₁₈N₂O₄).

4-Nitro-2-(trichloroacetyl)-1-methylimidazole (4b). A solution of 2-(trichloroacetyl)-1-methylimidazole (700 g, 3.1 mol) in 4 L of acetic anhydride was cooled to -40° C. Fuming nitric acid (500 ml), followed by conc. sulfuric acid (25 ml) was added over 2 h, and the reaction mixture was stirred for 12 h. The mixture was poured into a tray and allowed to crystallize over two days. The product was collected by vacuum filtration, washed (cold water, 1 × 100 ml), and dried *in vacuo* to yield **4b** as a white crystalline solid (472 g, 56% yield). TLC (1:1 hexanes/ethyl acetate) R_f 0.57 ^1H NMR (DMSO- d_6) δ 8.84 (s, 1 H), 4.04 (s, 3 H). FABMS m/e 270.932 (270.932 calcd for $\text{C}_6\text{H}_4\text{N}_3\text{O}_3$).

tert-Butyl 4-Nitro-1-methylimidazole-2-carboxylate (5b). To a mixture of 4-nitro-2-(trichloroacetyl)-1-methylimidazole (**3b**) (50 g, 0.18 mol) in 500 ml of *tert*-butyl alcohol was added sodium *tert*-butoxide (20 g) over 1 h. The reaction was refluxed under argon for 5 h, then quenched with 500 ml of water and extracted with chloroform (3 × 500 ml). The chloroform was removed *in vacuo* to provide **5b** (36.4 g, 84% yield). TLC (1:1 hexanes/ethyl acetate) R_f 0.59 ^1H NMR (DMSO- d_6) δ 8.62 (s, 1 H), 3.98 (s, 3 H), 1.58 (s, 9 H). FABMS m/e 227.091 (227.091 calcd for $\text{C}_9\text{H}_{13}\text{N}_3\text{O}_4$).

tert-Butyl 4-[(9-Fluorenylmethoxycarbonyl)amino]-1-methylimidazole-2-carboxylate (6b). *tert*-Butyl 4-Nitro-1-methylimidazole-2-carboxylate (**5b**) (9.2 g, 40 mmol) was dissolved in 100 ml of DMF. 10% Pd/C (1 g) in 20 ml of DMF was added. The mixture was vigorously stirred for 2 h under 500 psi of Hydrogen and then filtered through Celite. The Celite was rinsed with DMF (200 ml). Fmoc-Cl (10.7 g, 41 mmol) was added and the solution was stirred for 3 h. Water was added to the DMF and the mixture was extracted with diethyl ether. The ether was extracted 2 × water and 1 × brine, dried over sodium sulfate and concentrated by rotoevaporation to yield a crude yellow solid which was dissolved in a minimum amount of diethyl ether. Purified by column chromatography (3:1 hexanes/ethyl acetate, 5.9 g, 34.4% yield). TLC (1:1 hexanes/ethyl acetate) R_f 0.64 ^1H NMR (DMSO- d_6) δ 10.4 (s, 1 H), 7.9 (d, 2 H), 7.7 (d, 2

H), 7.4 (m, 2 H), 7.3 (m, 2 H), 7.2 (s, 1H),), 4.4 (d, 2 H, J = 6), 4.3 (t, 1 H, J = 6), 3.9 (s, 3 H), 1.6 (s, 9 H). FABMS m/e 419.185 (419.184 calcd for C₂₄H₂₅N₃O₄).

4-[(9-Fluorenylmethoxycarbonyl)amino]-1-methylimidazole-2-carboxylic Acid (7b). *tert*-Butyl 4-[(9-Fluorenylmethoxycarbonyl)amino]-1-methylimidazole-2-carboxylate (**6b**) (5.0 g, 11.9 mmol) was dissolved in 100 ml of dichloromethane and cooled to 0° C. TiCl₄ (25 ml, 1.0 M in dichloromethane) was added dropwise to the solution. The mixture was stirred for 1 h. Cold HCl (1 M, 500 ml) was added. The white precipitate was collected by vacuum filtration (3.7 g, 86% yield). ¹H NMR (DMSO-*d*₆) δ 12.4 (bs, 1 H), 10.4 (s, 1 H), 7.9 (d, 2 H), 7.7 (d, 2 H), 7.4 (m, 2 H), 7.3 (m, 2 H), 7.2 (s, 1H),), 4.4 (d, 2 H, J = 6), 4.3 (t, 1 H, J = 6), 3.9 (s, 3 H). FABMS m/e 364.131 (M+H) (364.122 calcd for C₂₀H₁₇N₃O₄).

Solid Phase Synthesis. Activation of Fmoc-Imidazole Acid. Activation of Imidazole-2-carboxylic acid, Fmoc-Pyrrole acid, Fmoc-Imidazole acid and γ-aminobutyric acid. The appropriate amino acid or acid (2 mmol) was dissolved in 2 mL of DMF. HBTU (720 mg, 1.9 mmol) was added followed by DIEA (1 mL) and the solution was lightly shaken for at least 5 min.

Machine-Assisted Protocols. Machine-assisted synthesis was performed on an ABI 430A synthesizer on a 0.69 mmol scale (800 mg resin; 0.86 mmol/g). Each cycle of amino acid addition involved deprotection with approximately 80% Piperidine/NMP for 3 min, draining the reaction vessel, and then deprotection for 17 min; two NMP washes; a DCM wash; an NMP wash; draining the reaction vessel; coupling for 1 h (8 h when coupling to imidazole), addition of dimethyl sulfoxide (DMSO)/NMP, coupling for 30 min; draining the reaction vessel; washing with DCM, taking a resin sample for evaluation of the progress of the synthesis for HPLC analysis; capping with acetic anhydride/DIEA in DCM for 6 min; and washing with DCM.

The ABI 430A synthesizer was left in the standard hardware configuration for NMP-HOBt protocols. Reagent position 1 was piperidine, reagent position 2 was not used, reagent position 3 was 70% ethanolamine/methanol, reagent position 4 was acetic anhydride, reagent position 5 was DMSO/NMP, reagent position 6 was methanol, reagent position 7 was DIEA, and reagent position 8 was NMP.

Fmoc-Im acid was added manually. Fmoc-imidazole acid (363 mg, 1 mmol) and HBTU (378 mg, 1 mmol) were combined in 2 mL of NMP. DIEA (1 ml) was then added, and the reaction mixture was allowed to stand for 5 min. At the initiation of the coupling cycle the synthesis was interrupted, the reaction vessel vented, and the activated monomer added directly to the reaction vessel. When manual addition was necessary, an empty synthesis cartridge was used.

Fmoc-Py acid (362 mg, 1 mmol) and aliphatic amino acids (2 mmol) were placed in a synthesis cartridge with 0.9 equiv. HBTU. NMP (3 ml) was added using a calibrated delivery loop from reagent bottle 8, followed by a calibrated delivery of 1 mL of DIEA from reagent bottle 7, and a 3 min mixing of the cartridge.

The activator cycle was written to transfer activated monomer directly from the cartridge to the concentrator vessel, bypassing the activator vessel. After transfer, 1 mL of DIEA was measured into the cartridge using a calibrated delivery loop, and the DIEA solution was combined with the activated monomer solution in the concentrator vessel. The activated ester in 2:1 DMF/DIEA was then transferred to the reaction vessel. All lines were emptied with argon before and after solution transfers.

ImPyPyPy- γ -PyPyPyPy- β -Dp (1). ImPyPyPy- γ -PyPyPyPy- β -Wang resin was prepared by machine-assisted synthesis protocols. A sample of resin (1 g, 0.47 mmol/g) was placed in a 20 mL glass scintillation vial. 4 mL of *N,N*-(dimethylamino) propylamine added, and the solution heated at 55° C for 18 h. Resin was removed by filtration through a disposable propylene filter, and 16 mL of water was added. The polyamide/amine mixture was purified directly by preparatory HPLC, and the appropriate fractions were

lyophilized to provide the trifluoroacetate salt of ImPyPyPy- γ -PyPyPyPy- β -Dp (240 mg, 38% recovery) as a white powder. MALDI-TOF-MS (monoisotopic), 1221.5 (1221.6 calc. for M+H).

ImPyPyPy- γ -ImPyPyPy- β -Dp (2). ImPyPyPy- γ -PyPyPyPy- β -Wang resin was prepared by machine-assisted synthesis protocols. A sample of resin (1 g, 0.47 mmol/g) was placed in a 20 mL glass scintillation vial. 4 mL of (dimethylamino)propylamine added, and the solution heated at 55 °C for 18 h. Resin was removed by filtration through a disposable propylene filter, and 16 mL of water was added. The polyamide/amine mixture was purified directly by preparatory HPLC, and the appropriate fractions were lyophilized to provide the trifluoroacetate salt of ImPyPyPy- γ -ImPyPyPy- β -Dp (57 mg, 9% recovery) as a white powder. MALDI-TOF-MS (monoisotopic), 1222.5 (1222.6 calc. for M+H).

Stepwise HPLC Analysis. A resin sample (ca. 4 mg) was placed in a 4 mL glass test tube. 200 μ L of (*N,N*-dimethylamino)propylamine was added, and the mixture was heated at 100° C for 5 min. The cleavage mixture was filtered and a 25 μ L sample analyzed by analytical HPLC at 254 nm.

Acknowledgment. We are grateful to the National Institute of Health for research support, the Ralph M. Parsons Foundation for a predoctoral fellowship for N.R.W., J. Edward Richter for an undergraduate fellowship to J.M.T. and the Howard Hughes Medical Institute for a predoctoral fellowship to E.E.B.

References

1. Dervan, P. B., and Burli, R. W. (1999) *Current Opinion in Chemical Biology* 3, 688-693.
2. Baird, E. E., and Dervan, P. B. (1996) *Journal of the American Chemical Society* 118, 6141-6146.
3. Vazquez, E., Caamano, A. M., Castedo, L., and Mascarenas, J. L. (1999) *Tetrahedron Letters* 40, 3621-3624.
4. Konig, B., and Rodel, M. (1999) *Synthetic Communications* 29, 943-949.
5. Nishiwaki, E., Tanaka, S., Lee, H., and Shibuya, M. (1988) *Heterocycles* 27, 1945-1952.
6. Wang, F. J., and Hauske, J. R. (1997) *Tetrahedron Letters* 38, 8651-8654.
7. Kiselyov, A. S., and Armstrong, R. W. (1997) *Tetrahedron Letters* 38, 6163-6166.
8. Lee, J., Murray, W. V., and Rivero, R. A. (1997) *Journal of Organic Chemistry* 62, 3874-3879.
9. Pei, Y. Z., Houghten, R. A., and Kiely, J. S. (1997) *Tetrahedron Letters* 38, 3349-3352.

CHAPTER 3

Solid Phase Synthesis of Six-ring Polyamides with Truncated C-Terminal Tails

The text of this chapter was taken in part from a publication co-authored with Jason Belitsky, Doan H. Nguyen & Prof. Peter B. Dervan.

(J. M. Belitsky, D. H. Nguyen, N. R. Wurtz & P. B. Dervan *Bioorganic & Medicinal Chemistry* submitted 2001)

Abstract

Pyrrole-imidazole polyamides are synthetic ligands that bind to predetermined sequences of DNA in the minor groove. Typically, solid phase methods using Boc monomers for synthesis have depended on Boc- β -Ala-PAM resin, which affords a β -alanine-Dp tail at the C-terminus, after cleavage with *N, N*-dimethylaminopropylamine (Dp). New solid phase strategies, including Boc chemistry on Kaiser oxime resin, have permitted the synthesis of polyamides with truncated C-terminal tails in high yield. Polyamides having carboxylic acid, ethyl, methyl, and primary amide tails show increased generality in the tail region relative to the standard β -Dp tail obtained from Boc- β -Ala-PAM resin synthesis. Polyamides with methyl amide tails rather than β -alanine show similar affinity relative to the standard β -Dp tail. The diminished preference for the A,T base pairs of truncated tails energetic will allow targeting of a greater number of DNA sequences by hairpin polyamides with high affinity.

Introduction

Hairpin polyamides containing *N*-methylpyrrole (Py), *N*-methylimidazole (Im), and *N*-methyl-3-hydroxypyrrole (Hp) residues bind specific predetermined sequences in the minor groove of DNA with affinities and specificities comparable to naturally occurring DNA-binding proteins (*1*). DNA recognition depends on the side-by-side amino acid pairings in the minor groove that stack the aromatic rings against each other and the walls of the groove allowing backbone amide hydrogens and the substituents at the 3-position of the Py, Im, and Hp residues to make specific contacts with the edges of the DNA base pairs. An Im/Py and Py/Im specifies for G•C and C•G, respectively. A Py/Py pair binds A•T and T•A and an Hp/Py pair discriminates T•A over A•T base pairs. These modular ring pairs may be considered the "core" of the polyamide recognition motif. For hairpins an alkyl amino acid, either γ -aminobutyric acid (GABA, γ), or the chiral, amine-functionalized derivative (*R*)-2,4-diaminobutyric acid (DABA, (*R*)^{H₂N} γ), serves as the covalent linker region (referred to as the "turn") between the N-terminal and C-terminal strands and is specific for A•T and T•A base pairs. The "tail" or C-terminal portion of a typical hairpin polyamide is related to the method of synthesis and comprises β -alanine (β) and *N,N*-dimethylaminopropylamine (Dp), which specify for A,T base pairs at the N-1 and N-2 positions (Figure 3.1) (*1*).

The β -Dp tail originated with the C \rightarrow N synthesis of polyamides by solid phase methods using Boc monomers on Boc- β -alanine-PAM-resin (2, 3). An alkyl primary amine such as Dp affords efficient cleavage from the solid support by aminolysis of the β -alanine ester linkage to the resin. The sequence requirements of the β -Dp tail for A,T versus G,C at the N-1 and N-2 positions (*4*) presumably results from a steric clash with

the exocyclic NH_2 of a G,C base pair on the floor of the minor groove of DNA (Figure 1). Replacement of the β -Dp tail with a shorter propanol amide, generated by the reductive cleavage of the β -alanine-PAM ester from the solid support (5), removes the preference for an A,T base pair at the N-2 position while maintaining the binding affinity for these sites, despite the loss of a positive charge (4). This suggests that the Dp tail interaction with the minor groove may be sterically unfavorable. A deeper understanding of the effect of incremental changes to the tail on the energetics of DNA binding would provide insight on the structural basis for A,T versus G,C specificity at the N-1 position.

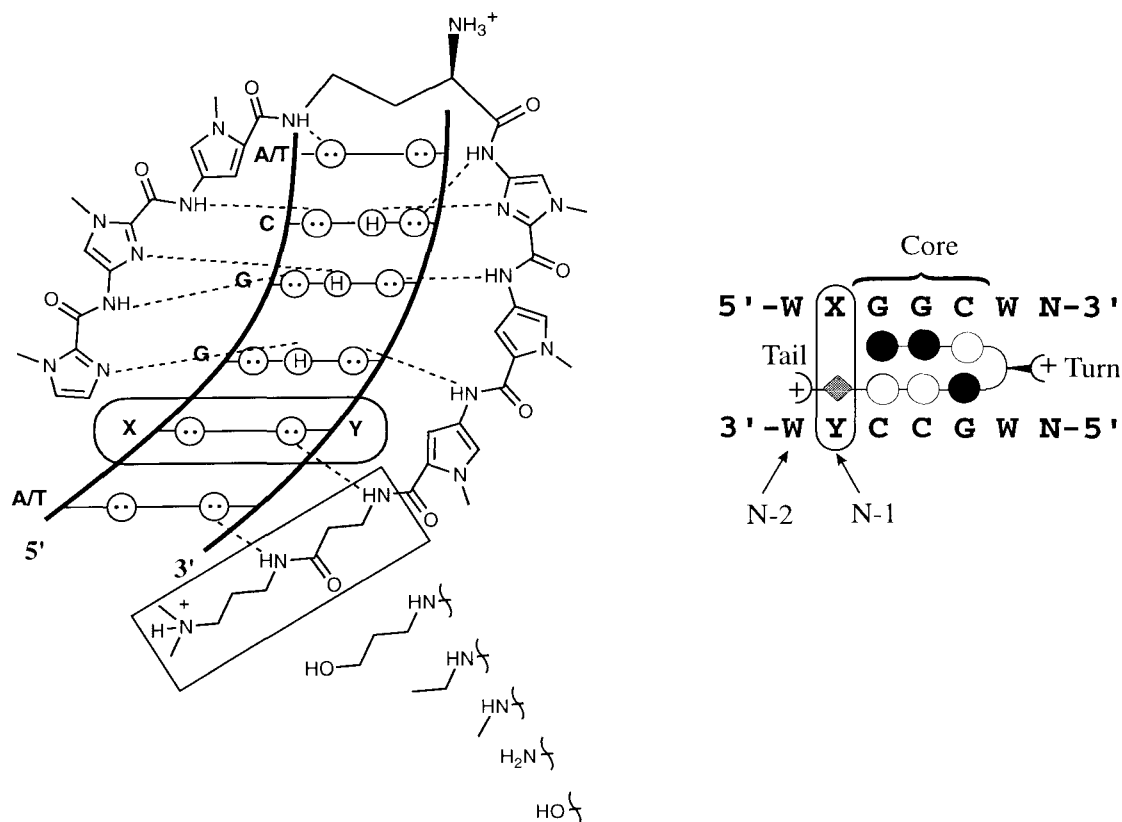


Figure 3.1. Binding model for polyamide **1** with DNA. (left) Circles with dots represent lone pairs of N3 of purines and O2 of pyrimidines. Circles containing an H represent the N2 hydrogens of guanine. Putative hydrogen bonds are illustrated by dotted lines. The series of truncated tails to be examined are denoted. (right) A ball and stick representation of polyamide **1** with DNA. Filled circles denote imidazole while open circles represent pyrrole. The diamond represents β -alanine. (*R*)-diaminobutyric acid ($(R)^{\text{H}_2\text{N}}\gamma$) and dimethylaminopropylamine (Dp) are depicted as a curved line and a plus sign, respectively. W signifies A or T, N represents any nucleotide.

In order to incrementally truncate the polyamide C-terminus, new solid phase methodology would be required since the propanol amide tail is the shortest C-terminus that can be generated from cleavage of Boc- β -Ala-PAM resin. Since the reductive cleavage to generate the propanol amide tail does not proceed reliably and results in poor overall recovery, it is generally not used. Therefore, for high affinity recognition hairpin polyamides are synthesized with β -Dp tails which require DNA binding sites to be composed of the sequence 5'-WW(N)_xW-3' (W = A,T; N = A,T,G,C; and x = the number of ring pairs in the core of the polyamide). While a large number of biologically relevant sequences have been targeted by the current generation of hairpin polyamides (6), the necessity for A,T base pairs flanking the polyamide core is a limitation. Promoters are known to have high G,C content associated with CpG islands (7), and many transcription factors bind G,C rich sequences (8). In order to effectively inhibit these transcription factors, polyamides that can readily target all base pairs at the tail positions with high affinity are desirable.

We describe here a new solid phase synthesis with a pyrrole unit directly linked to the Kaiser oxime resin as well as an adaptation of Fmoc solid phase synthesis to Rink resin. Six-ring polyamide hairpins with the common core sequence ImImPy-(R)^{H₂N} γ -ImPyPy-X (**3-6**) were synthesized where X represents four different C-terminal tails, ethyl amide, methyl amide, primary amide, and carboxylic acid. The DNA-binding affinity and specificity at the N-1 position was compared to the parent polyamides having either a β -Dp (**1**) or propanol amide tail (**2**).

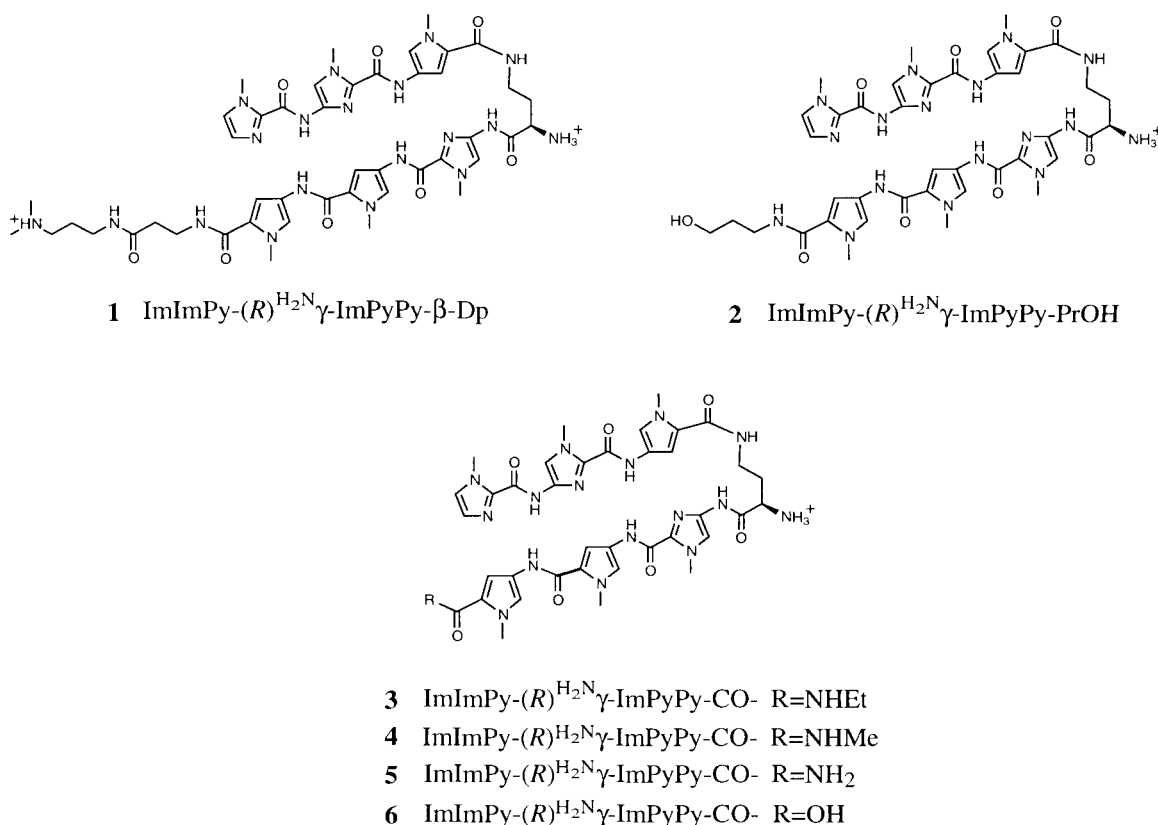


Figure 3.2. Structures of polyamides ImImPy-(*R*)^{H₂N}γ-ImPyPy-β-Dp (**1**), ImImPy-(*R*)^{H₂N}γ-ImPyPy-PrOH (**2**), ImImPy-(*R*)^{H₂N}γ-ImPyPy-CONHEt (**3**), ImImPy-(*R*)^{H₂N}γ-ImPyPy-CONHMe (**4**), ImImPy-(*R*)^{H₂N}γ-ImPyPy-CONH₂ (**5**) and ImImPy-(*R*)^{H₂N}γ-ImPyPy-CO₂H (**6**).

Results

Polyamide Synthesis. Polyamides ImImPy-(*R*)^{H₂N}γ-ImPyPy-β-Dp (**1**) and ImImPy-(*R*)^{H₂N}γ-ImPyPy-PrOH (**2**) were synthesized by solid phase methods on Boc-β-alanine-PAM-resin (**2**, **3**). The synthesis of Fmoc-protected Py (**7**) and Im (**8**) monomers and the production of β-Dp tail polyamides from Fmoc-β-alanine-Wang resin has recently been described (**9**). The Fmoc-protected monomers can also be used for the straightforward synthesis of polyamides with C-terminal primary amides. Commercially available Fmoc-Rink resin was deprotected with piperidine, and reacted with Fmoc-Py activated with HBTU and DIEA (**12**). Six standard cycles of deprotection and coupling

(9) yielded ImImPy-(R)^{BocHN γ} -ImPyPyCONH-Rink resin which was cleaved with TFA to yield **5** (Figure 3.3). Similar cleavage yields and purities were obtained compared to synthesis of the analogous polyamide with standard Boc protocols on Boc- β -Ala-PAM resin followed by aminolysis with Dp (3).

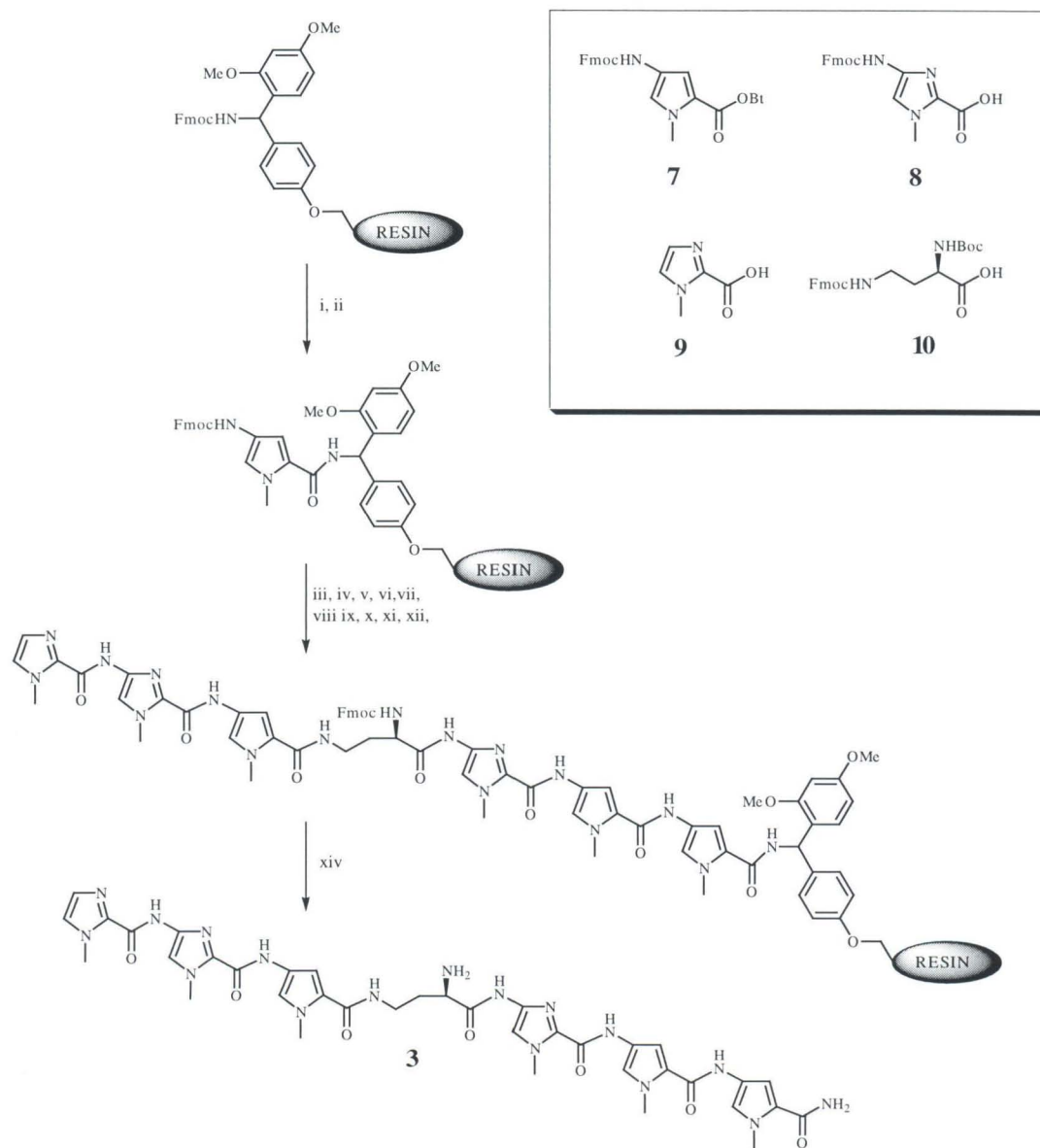


Figure 3.3. Synthesis of polyamide **3** from Rink resin using Fmoc-protected monomers: (i) 20% piperidine/DMF; (ii) **7**, HBTU, DIEA, NMP; (iii) 20% piperidine/DMF; (iv) **7**, HBTU, DIEA, NMP; (v) 20% piperidine/DMF; (vi) **8**, HBTU, DIEA, NMP; (vii) piperidine/DMF; (viii) **10**, HBTU, DIEA, NMP, 37 °C; (ix) piperidine/DMF; (x) **7**, HBTU, DIEA, NMP; (xi) piperidine/DMF; (xi) **8**, HBTU, DIEA, NMP; (xii) piperidine/DMF; (xiii) **9**, HBTU, DIEA, NMP; (xiv) TFA.

Four new polyamides **3-6** were synthesized by solid phase methods on the Kaiser oxime resin with Boc-pyrrole and Boc-imidazole monomers (Figure 3.4). Developed by Kaiser and DeGrado (10, 11), the oxime resin is a versatile polystyrene solid support that is amenable to Boc chemistry and allows for the synthesis of a variety of carboxylic acid derivatives by nucleophilic cleavage from the resin. Peptides can be cleaved from oxime resin to yield primary amide (11), alkyl amide (12), and carboxylic acid (13, 14) C-termini. The oxime linker is reported to be somewhat acid labile, and solutions of less than 25% TFA are recommended for the deprotection of Boc groups. Consequently, a 20% TFA/CH₂Cl₂ solution was used for 30 minutes to remove the Boc groups from pyrrole and aliphatic amines. However, it was found that Boc-imidazole residues could not be fully deprotected under these conditions. Longer deprotection times resulted in increasing degradation of the resin bound polyamide. Boc groups could be efficiently removed from imidazoles in 30 minutes with a 50% TFA/CH₂Cl₂ solution. Remarkably, no premature cleavage of the oxime linker was observed at this higher concentration of acid. Attachment of the first pyrrole unit results in a bulky aromatic oxime ester that apparently stabilizes the linker relative to the aliphatic oxime ester present when conventional amino acids are used.

Polyamides **3-6** were generated from the common intermediate, ImImPy-(*R*)^{FmocHN}-γ-ImPyPy-Oxime resin **15** (Figure 3.4). Commercially available oxime resin was allowed to react directly with Boc-Py-OBt (**11**) in *N*-methylpyrrolidone (NMP) and diisopropylethylamine (DIEA) overnight at room temperature. The subsequent deprotections and couplings were carried out in a stepwise manner using the previously outlined deprotection conditions. Monomers **12-14** were activated with HBTU and

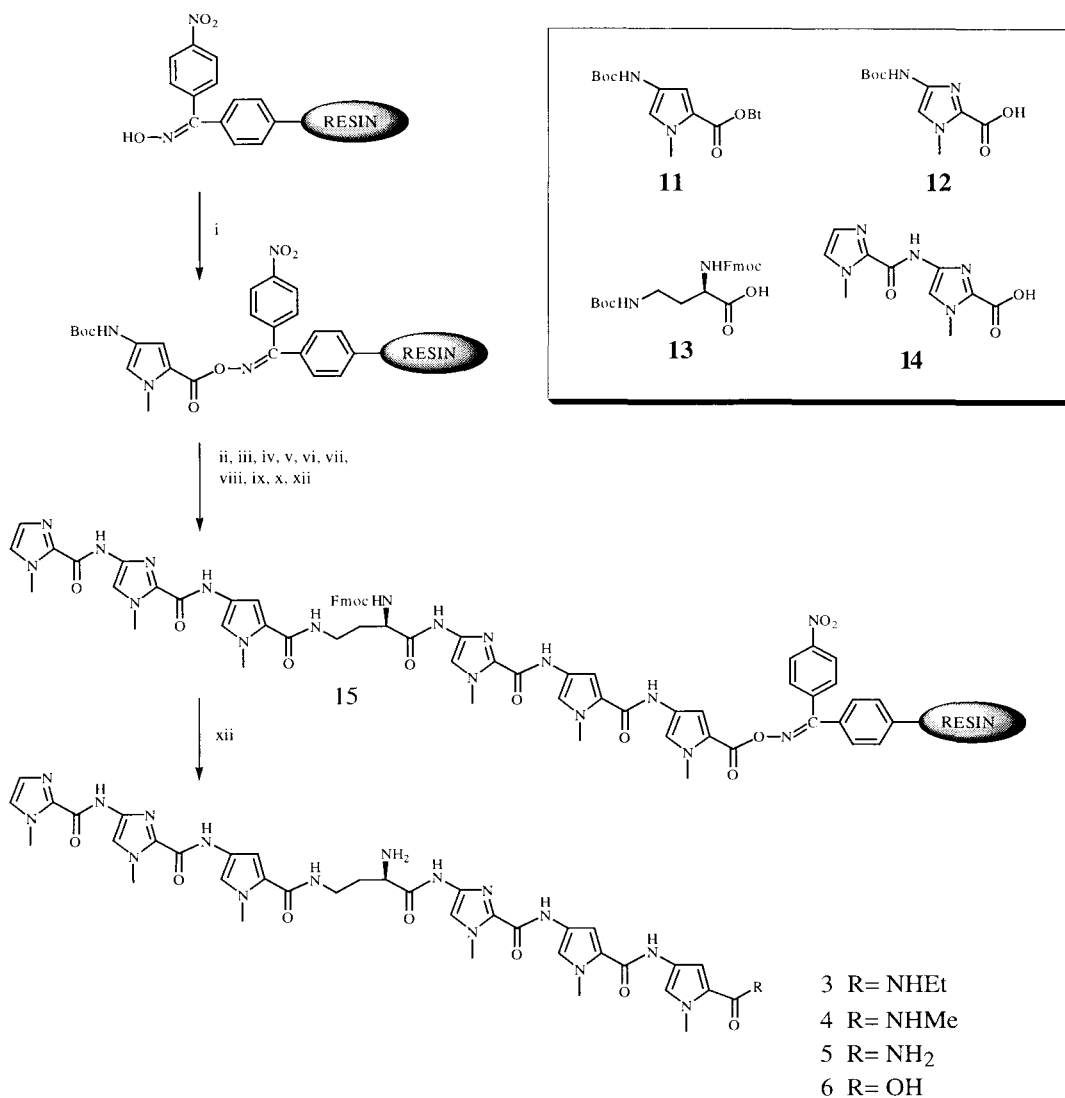


Figure 3.4. Synthesis of polyamide **3-6** from oxime resin using Boc-protected monomers: (i) **11**, DIEA, NMP; (ii) 20% TFA/DCM; (iii) **11**, DIEA, NMP; (v) 20% TFA/DCM; (v) **12**, HBTU, DIEA, NMP; (vi) 50% TFA/DCM; (vii) **13**, HBTU, DIEA, NMP, 37 °C; (viii) 20% TFA/DCM; (ix) **11**, DIEA, NMP; (x) 20% TFA/DCM; (xi) **14**, HBTU, DIEA, NMP; (xii) a) R = NHEt, 2.0 M CH₃CH₂NH₂/THF, DCM, 37°C; b) R = NHMe, 2.0 M CH₃NH₂/THF, DCM, 37 °C; c) R = NH₂, NH₃/THF, DCM, DBU, 37°C; or d) R = OH, H₂O, DMF, DBU, 37 °C.

coupled in NMP/DIEA for 1.5-8 hrs at either room temperature or 37°C. All cleavages of resin **15** required a co-solvent (CH₂Cl₂ or DMF) to swell the resin and/or solubilize the polyamide. Based on literature methods (11), cleavage of resin **15** with a saturated solution of ammonia in THF/CH₂Cl₂ yielded polyamide **5**. However, even after more than 60 hrs at 37°C only 60% cleavage was observed. Addition of a large excess of

DBU, which had been used as a transacylation catalyst to generate peptide acids and esters from oxime resin previously (13), resulted in quantitative cleavage of the resin to generate polyamide **5**. The C-terminal carboxylic acid polyamide **6** was generated in a similar manner with 1:1 H₂O:DMF, 30 eq. DBU for 60 hrs at 37 °C. This method was superior to acidic, basic, or reductive cleavage conditions. It is interesting to note that the standard transacylation catalyst, DMAP, was ineffective as a replacement for DBU in these cleavage reactions. Both polyamides **3** and **4** were generated in high yield from resin **15** with a 1.0 M solution in THF/CH₂Cl₂ of ethylamine or methylamine, respectively, overnight at 37°C. The oxime resin is a versatile way to generate modified polyamides under mild conditions in good yield while remaining amenable to Boc chemistry.

Quantitative DNase I Footprinting. It has been shown previously that elimination of Dp in the β-Dp tail in polyamide **1** diminishes the requirement for A,T preference at the N-2 position (4). While empirical observations (15) have suggested that the propanol amide tail, like the β-Dp tail (**2**) is A,T specific at the N-1 position, this had not been investigated in detail. Quantitative DNase I footprinting titrations (16) were used to determine the equilibrium association constants (K_a) of polyamides **2-6** for four sites on a restriction fragment corresponding to the sequences 5'-AXGGCTA-3'(where X = A,T,G,C) at the N-1 position (Figure 4A-D, Table 1). Conversion from β-Dp (**1**) to propanol amide (**2**) results in similar affinity and specificity. This retention of affinity upon removal of the positively charged Dp moiety has previously been observed (17, 18). Presumably, the steric clash of propanol amide versus β-Dp with the exocyclic NH₂ of G at N-1 is similar.

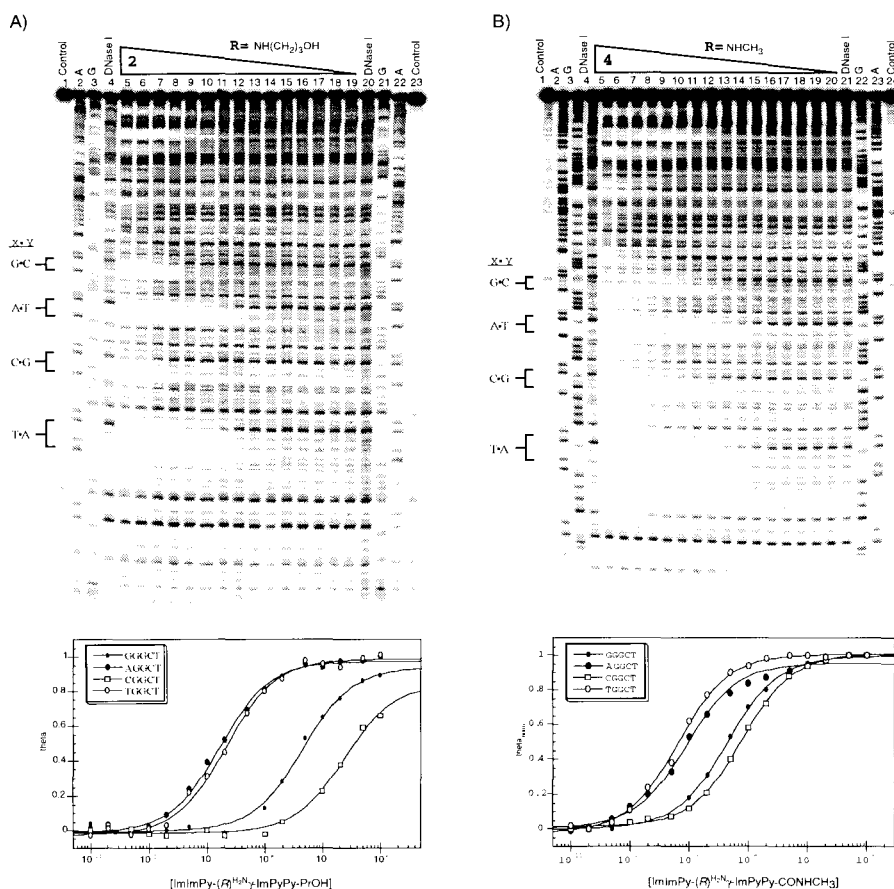


Figure 4 A&B. Quantitative DNase I footprint titration experiment with ImImPy-(R)^{H2N}-ImPyPyPrOH (**2**) and ImImPy-(R)^{H2N}-ImPyPyCONHCH₃ (**4**) on the 3'-³²P-labeled 286-bp restriction fragment pSES-TL1(4): lane 1 & 23, intact DNA; lane 2 & 22, A reaction; lane 3 & 21, G reaction; lane 4 & 20 DNase I standard; lanes 5-19 DNase I digestion products in the presence of 1 μ M, 500 nM, 200 nM, 100 nM, 20 nM, 10 nM, 5 nM, 2 nM, 1 nM, 500 pM, 200 pM, 100 pM, 50 pM, 20 pM, 10 pM polyamide respectively. All reactions contain 15 kbp restriction fragment, 10 mM Tris•HCl (pH 7.0), 10 mM KCl, 10 mM MgCl₂, and 5 mM CaCl₂. Data was obtained for the four binding sites indicated at the left of the gel, 5'-GGGCT-3', 5'-AGGCT-3', 5'-CCGCT-3', and 5'-TGGCT-3', and is shown in the isotherm plot below. θ_{norm} points were obtained using storage phosphor autoradiography and processed by standard methods. Each data point shows the average value obtained from three footprinting experiments. The solid curves are best-fit Langmuir binding titration isotherms obtained from nonlinear least squares algorithm where $n = 1$ as previously described. (2)

Polyamides **3-5** afford similar DNA binding affinity versus the parent β -Dp while diminishing the sequence preference for A,T versus G,C at N-1. Polyamide **3** (ethyl amide) maintains a strong preference for A,T versus G,C, while polyamide **4** (methyl

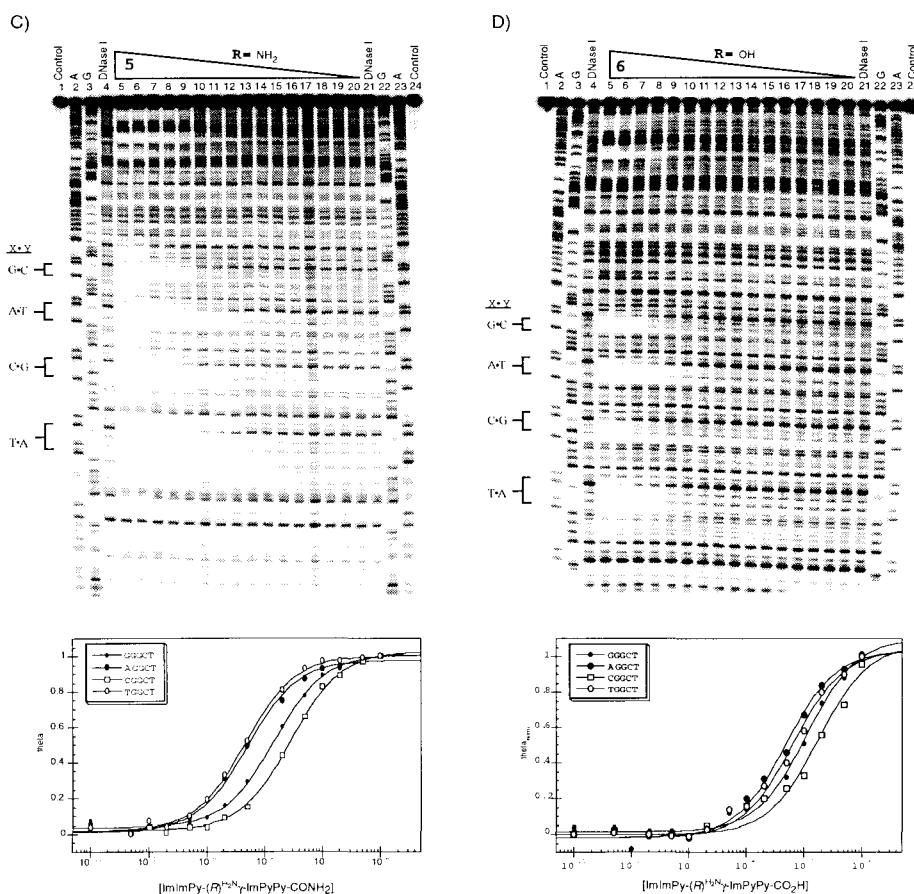
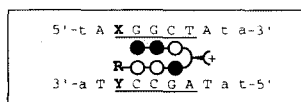


Figure 4 C&D. Quantitative DNase I footprint titration experiment with ImImPy-(R)^{H₂N}γ-ImPyPyCONH₂ (**5**) and ImImPy-(R)^{H₂N}γ-ImPyPyCO₂H (**6**) on the 3'-³²P-labeled 286-bp restriction fragment pSES-TL1(4): lane 1 & 24, intact DNA; lane 2 & 23, A reaction; lane 3 & 22, G reaction; lane 4 & 21 DNase I standard; lanes 5-20 DNase I digestion products in the presence of 1 μM, 500 nM, 200 nM, 100 nM, 50 nM, 20 nM, 10 nM, 5 nM, 2 nM, 1 nM, 500 pM, 200 pM, 100 pM, 50 pM, 20 pM, 10 pM polyamide respectively. Conditions and data analysis as outlined in Figure 4 A&B.

amide) less so. Polyamide **5** (primary amide) shows a loss in affinity relative to **1-4**, and little sequence discrimination between the four Watson-Crick base pairs. Polyamide **6** (carboxylic acid) reveals poor affinity and little sequence discrimination at the N-1 position (Table I). The decreased affinity of polyamides **5** and **6** suggests that the increased Van der Waal contacts with the walls of the minor groove, made by polyamides **1-4**, may be important.



R=	-β-Dp (1)	-NH(CH ₂) ₃ OH (2)	-NHCH ₂ CH ₃ (3)	-NHCH ₃ (4)	-NH ₂ (5)	-OH (6)
X•Y						
A•T	1.7 X 10 ⁹ (±0.5)	5.9 X 10 ⁸ (±0.4)	1.1 X 10 ⁹ (±0.1)	1.1 X 10 ⁹ (±0.1)	1.9 X 10 ⁸ (±0.2)	2.1 X 10 ⁷ (±0.1)
T•A	1.6 X 10 ⁹ (±0.2)	4.7 X 10 ⁸ (±0.4)	9.2 X 10 ⁸ (±0.3)	1.4 X 10 ⁹ (±0.3)	2.1 X 10 ⁸ (±0.2)	1.2 X 10 ⁷ (±0.2)
G•C	5.8 X 10 ⁷ (±1.4)	2.1 X 10 ⁷ (±0.3)	8.2 X 10 ⁷ (±1.3)	2.4 X 10 ⁸ (±0.1)	6.8 X 10 ⁷ (±0.5)	1.1 X 10 ⁷ (±0.2)
C•G	3.2 X 10 ⁷ (±0.3)	6.2 X 10 ⁶ (±0.7)	3.0 X 10 ⁷ (±0.5)	1.1 X 10 ⁸ (±0.2)	3.5 X 10 ⁷ (±0.2)	4.9 X 10 ⁶ (±0.2)

Table 3.1. Equilibrium association constants (M^{-1}) for polyamides **1-6** with restriction fragment pSES-TL1. (4) Values reported are the mean values from at least three DNase I footprinting titration experiments with the standard deviation for each data set in parentheses. Equilibrium association constants for polyamide **1** at the N-1 position were obtained from Swalley *et al.* (4).

Discussion

In sequences with an A•T or T•A at the N-1 position, structural evidence (19-24) suggests that the C-terminal pyrrole secondary amide makes a hydrogen bond with the N3 of adenine and/or the O2 of thymine (Figure 1). However, hydrogen bonds could also be made to the N3 of guanine and/or the O2 of cytosine. Assuming these hydrogen bonds are similar energetically, the discrimination of A,T versus G,C is likely due to steric reasons. Thus, the methylene groups of the propanol amide and β-Dp tails clash with the exocyclic amine of guanine and by default incur a lower energetic penalty at A,T. The influence of the exocyclic amine of guanine on specificity can be inferred from the observation that the sequence XY = C•G was the lowest affinity site for all of the polyamides examined in this study. At this site the guanine base is on the bottom strand of the DNA proximal to the C-terminus of the hairpin where the steric clash between the tail and the exocyclic NH₂ of guanine should be most severe.

Polyamides **3** and **4** with the ethyl and methyl amide tails provide a way to investigate whether specificity could be assigned to a particular methylene group of the propanol amide tail. The specific hydrogen bond made by the amide of the C-terminal pyrrole would point the first methylene group away from the floor of the minor groove, such that the energetic penalty and hence specificity might be attributable to the second methylene group. However, the methyl amide did exhibit about 8-fold preference for binding A,T or G,C. The ethyl amide tail disfavors both G•C and C•G sites by about 3-fold more than the methyl amide tail. Clearly, both alkyl groups contribute to A,T specificity, and even greater steric bulk beyond the second carbon is likely necessary to account for the greater than 50-fold specificity of the β -Dp and propanol amide tails. The high affinity of polyamides **3** and **4** suggest an energetic balance between increased Van der Waals contacts with the walls of the minor groove relative to the primary amide tail (**5**) and decreased steric clashing relative to the β -Dp tail (**1**).

Polyamides with C-terminal carboxylic acid tails have been synthesized as intermediates to cyclic polyamides but their DNA binding properties have not been investigated (25, 26). The carboxylic acid tail polyamide (**6**) binds all four sites with low affinity and poor sequence discrimination. To act as a hydrogen bond donor in analogy to the amides, the carboxylic acid would have to be protonated in the minor groove. The acid is likely deprotonated in the minor groove and the 10-fold lower affinity compared to the primary amide tail may be due to loss of this hydrogen bond. Additionally, in water polyamide **6** is expected to be a zwitterion with a net charge of zero, which one would expect to have a lower affinity for DNA than the positively charged polyamides **1-5**.

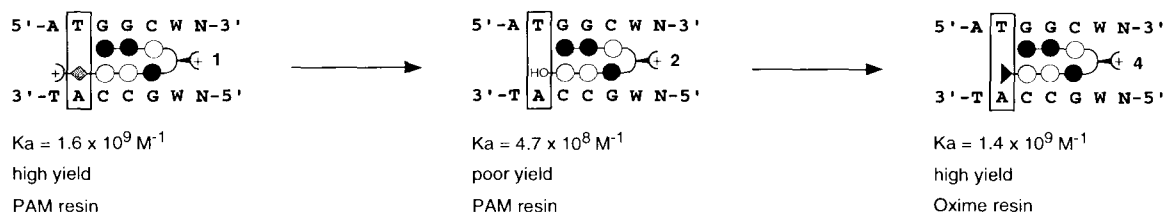


Figure 3.5. Summary of end effects by truncated polyamides. For the same DNA sequence a 20-fold increase is observed moving from β -Dp polyamide **1** to propanol amide tail polyamide **2**. A further 3-fold increase is observed in truncation of the propanol amide to the methyl amide tail polyamide **4**.

Conclusion

In summary, hairpin polyamides with incrementally truncated C-terminal tails can be synthesized using Boc chemistry, in high yield, on oxime resin by solid phase methods (Figure 3.5). The polyamides are lower molecular weight than the corresponding β -Dp polyamide and bind all four base pairs at the N-1 position with high affinity, which will enable targeting of new biologically relevant sequences by hairpin polyamides. If, at the N-1 position, one desires specificity for A,T > G,C while maintaining high affinity, then the propanol amide or the ethyl amide is the end group of choice. If at the N-1 position, one prefers less sequence preference with high affinity, then the methyl amide is the best choice. It will be interesting to study the influence of truncated tails on cellular uptake and nuclear localization of hairpin polyamides.

Experimental

Materials. Polyamides **1** and **2**, and the restriction fragment pSES-TL1 have been previously described (4). DNase I footprinting titrations on the ^{32}P -labeled restriction fragment pSES-TL1 were performed as previously described (4). 2.0 M methylamine in THF, 2.0 M ethylamine in THF, DBU and anhydrous ammonia gas were

purchased from Aldrich. Oxime resin was purchased from Nova Biochem. All other synthetic and footprinting reagents were as previously described (3, 16, 18). ^1H NMR were recorded on a Varian Mercury 300 instrument. UV spectra were measured on a Beckman Coulter DU 7400 diode array spectrophotometer. Autoradiography was performed with a Molecular Dynamics Typhoon Phosphorimager. Matrix-assisted, laser desorption/ionization time of flight (MALDI-TOF) mass spectrometry was carried out at the Peptide and Protein Microanalytical Facility at the California Institute of Technology. HPLC analysis was performed on a Beckman Gold system using a RAINEN C_{18} , Microsorb MV, 5 μm , 300 x 4.6 mm reversed-phase column in 0.1% (w/v) TFA with CH_3CN as eluent and a flow rate of 1.0 mL/min, gradient elution 1.25% $\text{CH}_3\text{CN}/\text{min}$. Preparatory HPLC was carried out on a Beckman HPLC using a Waters DeltaPak 25 x 100 mm, 100 μm C_{18} column, 0.1% (w/v) TFA, 0.56% $\text{CH}_3\text{CN}/\text{min}$. 18M Ω water was obtained from a Millipore MilliQ water purification system, and all buffers were 0.2 μm filtered. Reagent-grade chemicals were used unless otherwise stated.

ImImPy-(R) $^{\text{H}_2\text{N}}$ γ -ImPyPyCONHEt (3). ImImPy-(R) $^{\text{BocHN}}$ γ -ImPyPyCO-Oxime resin (**15**) was generated by manual solid phase synthesis from oxime resin (1 g, 0.48 mmol/g) using previously described Boc-protected monomers (3). Boc-Py-OBt (**11**) (358 mg, 1 mmol) was dissolved in 2 ml of NMP and added to 1 g of oxime resin followed by 1 ml of DIEA. The coupling was allowed to proceed overnight at room temperature. The resin was then acetylated with 3 ml of acetic anhydride (Ac_2O), 4 ml of NMP and 1 ml of DIEA for 30 minutes at room temperature. The Boc group was removed upon treatment with 20% TFA/DCM for 30 minutes. The second pyrrole residue was coupled in the same fashion as the first but complete coupling could be achieved in 2 hours at room

temperature followed by the acetylation and deprotection steps outlined above. Boc-Im-OH (**12**) (241 mg, 1 mmol) was dissolved in 2 ml of NMP to which 360 mg (1 mmol) of HBTU and 1 ml of DIEA was added for activation of this monomer. Coupling was allowed to proceed for 1.5 hours at room temperature followed by acetylation. The Boc-Im residue was deprotected using a 50% TFA/DCM solution for 30 minutes at room temperature. The next residue, α -Fmoc- γ -Boc-(*R*)-diaminobutyric acid (DABA) (**14**) (660 mg, 1.5 mmol) was activated with HBTU (540 mg, 1.5 mmol) in 2 ml of NMP and 1 ml of DIEA. Coupling of this residue onto the resin took 2 hours at 37 °C. After acetylation and treatment with 20% TFA/DCM for 30 minutes, the next Boc-Py-OBt was attached in exactly the same manner as the second residue. The terminal two imidazoles were added as an ImIm-OH dimer (**3**). 249 mg (1 mmol) of the ImIm-OH dimer (**14**) was activated with HBTU (360 mg, 1 mmol) in 2 ml of NMP and 1 ml of DIEA and allowed to couple overnight at room temperature. It should be noted that the progress of the stepwise couplings were all monitored by analytical HPLC. The resin was washed thoroughly with DMF, DCM, MeOH and Et₂O then dried *in vacuo*. A 75 mg sample of dried resin was suspended in 2 ml of CH₂Cl₂ to which was added 2 ml of 2.0 M ethylamine in THF. This cleavage mixture was placed in a 37 °C oven and allowed to stand overnight in a sealed scintillation vial. The resin was filtered, the eluent concentrated *in vacuo*, then purified by reverse phase HPLC. ImImPy-(*R*)^{H₂N} γ -ImPyPyCONHCH₃ (1.9 mg, 2.2 μ mol, 10.6% recovery) was recovered upon lyophilization of the appropriate fractions as a white powder; UV (H₂O) λ_{max} 310 (52140); ¹H NMR (DMSO-*d*₆) δ 11.02 (s, 1H), 10.36 (s, 1H), 10.10 (s, 1H), 9.89 (s, 1H), 9.71 (s, 1H), 8.20 (m, 1H), 7.98 (m, 1H), 7.56 (s, 1H), 7.52 (s, 1H), 7.45 (d, 1H, *J* = 0.9 Hz), 7.25

(d, 1H, $J = 1.5$ Hz), 7.23 (d, 1H, $J = 1.5$ Hz), 7.15 (d, 1H, $J = 1.8$ Hz), 7.11 (d, 1H, $J = 1.8$ Hz), 7.06 (d, 1H, $J = 0.9$ Hz), 7.03 (d, 1H, $J = 1.8$ Hz), 6.83 (d, 1H, $J = 1.8$ Hz), 3.99 (s, 3H), 3.98 (s, 3H), 3.96 (s, 3H), 3.83 (s, 3H), 3.80 (s, 3H), 3.78 (s, 3H), 3.17 (dd, 2H, $J = 6.6$ Hz), 2.42 (m, 2H), 2.26 (m, 1H), 1.99 (m, 2H), 1.62 (m, 1H), 1.05 (t, 3H $J = 6.6$ Hz), MALDI-TOF-MS 866.5 (866.39 calc for $[M+H]^+ C_{39}H_{48}N_{17}O_7$).

ImImPy-(R)^{H₂N} γ -ImPyPyCONH₂ (3) (Fmoc/Rink Resin Route). ImImPy-(R)^{BocHN} γ -ImPyPyCONH-Rink resin was generated by manual solid phase synthesis from Fmoc-Rink resin (0.6 mmol/g) using Fmoc-protected monomers as previously described (3). A sample of the polyamide resin (30 mg) was cleaved by treatment of 20% TFA in DCM at room temperature for 2 hours. After filtration the cleavage solution was concentrated *in vacuo* and purified by reverse phase HPLC as previously described (3). ImImPy-(R)^{H₂N} γ -ImPyPyCONH₂ (630 μ g, 750 nmol, 4.2% recovery) was recovered upon lyophilization of the appropriate fractions as a white powder. Spectroscopic analysis confirmed this product was identical to **3** produced by the Oxime route.

ImImPy-(R)^{H₂N} γ -ImPyPyCONHMe (4). A 35 mg sample of dried resin **15** was suspended in 2 ml of CH₂Cl₂ to which was added 2 ml of 2.0 M methylamine in THF. This cleavage mixture was placed in a 37°C oven and allowed to stand overnight in a sealed scintillation vial. The resin was filtered, the eluant concentrated *in vacuo*, then purified by reverse phase HPLC. ImImPy-(R)^{H₂N} γ -ImPyPyCONHCH₃ (850 μ g, 996 nmol, 8.9% recovery) was recovered upon lyophilization of the appropriate fractions as a white powder; UV (H₂O) λ_{\max} 310 (52140); ¹H NMR (DMSO-*d*₆) δ 11.01 (s, 1H), 10.35 (s, 1H), 10.09 (s, 1H), 9.90 (s, 1H), 9.70 (s, 1H), 8.19 (m, 1H), 7.92(m, 1H), 7.56 (s, 1H), 7.51 (s,

1H), 7.45 (d, 1H, $J = 0.9$ Hz), 7.25 (d, 1H, $J = 1.8$ Hz), 7.23 (d, 1H, $J = 1.8$ Hz), 7.16 (d, 1H, $J = 1.5$ Hz), 7.10 (d, 1H, $J = 1.5$ Hz), 7.05 (d, 1H, $J = 0.9$ Hz), 7.03 (d, 1H, $J = 1.8$ Hz), 6.79 (d, 1H, $J = 1.8$ Hz), 3.99 (s, 3H), 3.98 (s, 3H), 3.96 (s, 3H), 3.82 (s, 3H), 3.79 (s, 3H), 3.77 (s, 3H), 2.66 (m, 1H), 2.41 (m, 2H), 2.25 (m, 1H), 1.98 (m, 2H), 1.62 (m, 2H), MALDI-TOF-MS 852.5 (852.38 calc for $[M+H]^+ C_{38}H_{46}N_{17}O_7$).

ImImPy-(R)^{H₂N}γ-ImPyPyCONH₂ (5). A 30 mg sample of dried resin **15** was placed into a pressure tolerant screw cap test tube, suspended in 4 ml of dry THF and cooled for 20 minutes in a -10°C ice/brine bath. Anhydrous NH₃ gas was bubble into the suspension at a steady rate for 30 minutes to generate a saturated ammonia solution. 1.5 ml of a 10%v/v solution of DBU in CH₂Cl₂, cooled to -20°C was quickly added to the cleavage mixture and the tube immediately sealed. The cleavage was allowed to proceed in a shaker at 37°C for 72 hours. The mixture is filtered and concentrated *in vacuo*, then purified by reverse phase HPLC. ImImPy-(R)^{H₂N}γ-ImPyPyCONH₂ (330 μg, 390 nmol, 4.1% recovery) was recovered upon lyophilization of the appropriate fractions as a white powder; UV (H₂O) λ_{max} 310 (52140); ¹H NMR (DMSO-*d*₆) δ 10.65 (s, 1H), 10.32 (s, 1H), 10.20 (s, 1H), 9.90 (s, 1H), 9.75 (s, 1H), 8.09 (bs, 1H), 7.55 (s, 3H), 7.47 (s, 1H), 7.44 (s, 1H), 7.25 (d, 1H, $J = 2.1$ Hz), 7.21 (d, 1H, $J = 2.1$ Hz), 7.19 (bs 2H), 7.11 (d, 1H, $J = 1.8$ Hz), 7.05 (d, 1H, $J = 1.2$ Hz), 6.98 (d, 1H, $J = 1.2$ Hz), 6.83 (d, 1H, $J = 1.5$ Hz), 6.65 (d, 1H, $J = 1.2$ Hz), 6.53 (d, 1H, $J = 2.4$ Hz), 3.99 (s, 3H), 3.98 (s, 3H), 3.96 (s, 3H), 3.82 (s, 3H), 3.79 (s, 1H), 3.77 (s, 3H), 2.97 (m, 1H), 1.96 (m, 2H), 1.43 (m, 2H), MALDI-TOF-MS 838.4 (838.35 calc for $[M+H]^+ C_{37}H_{44}N_{17}O_7$).

ImImPy-(R)^{H₂N}γ-ImPyPyCO₂H (6). A 35 mg sample of dried resin **15** was cleaved by treatment with 1 ml of H₂O, 1 ml DMF, and 75 μl DBU (0.48 mmol) at 37°C

for 60 hours. After filtration the cleavage solution was concentrated *in vacuo* and purified by reverse phase HPLC. ImImPy-(R)^{H₂N}γ-ImPyPyCO₂H (294 μg, 350 nmol, 3.1% recovery) was recovered upon lyophilization of the appropriate fractions as a white powder; UV (H₂O) λ_{max} 310 (52140); ¹H NMR (DMSO-*d*₆) δ 12.67 (bs, 1H), 11.53 (s, 1H), 10.88 (s, 1H), 10.62 (s, 1H), 10.43 (s, 1H), 10.22 (s, 1H), 8.83 (m, 1H), 8.07 (s, 1H), 8.03 (s, 1H), 7.96 (d, 1H, *J* = 0.9 Hz), 7.93 (d, 1H, *J* = 1.8 Hz), 7.76 (d, 1H, *J* = 0.9 Hz), 7.75 (d, 1H, *J* = 1.8 Hz), 7.64 (d, 1H, *J* = 1.2 Hz), 7.57 (d, 1H, *J* = 1.2 Hz), 7.54 (d, 1H, *J* = 1.8 Hz), 7.34 (d, 1H, *J* = 1.8 Hz), 4.50 (s, 3H), 4.50 (s, 3H), 4.47 (s, 3H), 4.34 (s, 3H), 4.31 (s, 3H), 4.31 (s, 3H), 2.50 (m, 2H), 2.11 (m, 2H), 2.06 (m, 3H), MALDI-TOF-MS 839.4 (839.34 calc for [M+H] C₃₇H₄₃N₁₆O₈⁺).

Quantitative DNase I Footprinting. As previously reported (16).

Acknowledgments. We are grateful to the National Institutes of Health GM27681 for research support, the Ralph M. Parsons Foundation for a predoctoral fellowship to J.M.B., the Natural Sciences and Engineering Research Council of Canada for a postgraduate scholarship to D.H.N., and Bristol-Myers Squibb for a predoctoral fellowship to N.R.W.

References

1. Dervan, P. B., and Bürli, R. W. (1999) *Current Opinions in Chemical Biology* 3, 688-693.
2. Parks, M. E., Baird, E. E., and Dervan, P. B. (1996) *Journal of the American Chemical Society* 118, 6147-6152.
3. Baird, E. E., and Dervan, P. B. (1996) *Journal of the American Chemical Society* 118, 6141-6146.
4. Swalley, S. E., Baird, E. E., and Dervan, P. B. (1999) *Journal of the American Chemical Society* 121, 1113-1120.
5. Trauger, J. W., Baird, E. E., and Dervan, P. B. (1998) *Angewandte Chemie-International Edition* 37, 1421-1423.
6. Gottesfeld, J. M., Turner, J. M., and Dervan, P. B. (2000) *Gene Expression* 9, 77-91.
7. Ioshikhes, I. P., and Zhang, M. Q. (2000) *Nature Genetics* 26, 61-63.
8. Pabo, C. O., and Sauer, R. T. (1992) *Annual Review of Biochemistry* 61, 1053-1095.
9. Wurtz, N. R., Turner, J. M., Baird, E. E., and Dervan, P. B. (2001) *Organic Letters* 3, 1201-1203.
10. DeGrado, W. F. and Kaiser, E.T. (1980) *Journal of Organic Chemistry* 45, 1295-1300.
11. DeGrado, W. F. and Kaiser, E.T. (1982) *Journal of Organic Chemistry* 47, 3258-3261.

12. Voyer, N., Lavoie, A., Pinette, M. and Bernier, J. (1994) *Tetrahedron Letters* 35, 355-358.
13. Pichette, A., Voyer, N., Larouche, R. and Meillon, J-C. (1997) *Tetrahedron Letters* 38, 1279-1282.
14. Nakagawa, S. H., Kaiser, E.T. (1983) *Journal of Organic Chemistry* 48, 678-685.
15. Dervan, P. B. *Unpublished Results*.
16. Trauger, J. W. and Dervan, P.B. (2001) *Methods in Enzymology* 340, 450-466.
17. Bremer, R. E., Wurtz, N.R., Szewczyk, J.W. and Dervan, P.B. (2001) *Bioorganic & Medicinal Chemistry* 9, 2093-2103.
18. Herman, D. M., Baird, E. E., and Dervan, P. B. (1998) *Journal of the American Chemical Society* 120, 1382-1391.
19. Bailly, C., and Chaires, J. B. (1998) *Bioconjugate Chemistry* 9, 513-538.
20. deClairac, R. P. L., Geierstanger, B. H., Mrksich, M., Dervan, P. B., and Wemmer, D. E. (1997) *Journal of the American Chemical Society* 119, 7909-7916.
21. deClairac, R. P. L., Seel, C. J., Geierstanger, B. H., Mrksich, M., Baird, E. E., Dervan, P. B., and Wemmer, D. E. (1999) *Journal of the American Chemical Society* 121, 2956-2964.
22. Kielkopf, C. L., Baird, E. E., Dervan, P. D., and Rees, D. C. (1998) *Nature Structural Biology* 5, 104-109.
23. Kielkopf, C. L., Bremer, R. E., White, S., Szewczyk, J. W., Turner, J. M., Baird, E. E., Dervan, P. B., and Rees, D. C. (2000) *Journal of Molecular Biology* 295, 557-567.

24. Kielkopf, C. L., White, S., Szewczyk, J. W., Turner, J. M., Baird, E. E., Dervan, P. B., and Rees, D. C. (1998) *Science* 282, 111-115.
25. Cho, J., Parks, M. E., and Dervan, P. B. (1995) *Proceedings of the National Academy of Sciences, USA* 92, 10389-10392.
26. Herman, D. M., Turner, J. M., Baird, E. E., and Dervan, P. B. (1999) *Journal of the American Chemical Society* 121, 1121-1129.

CHAPTER 4

Inhibition of Major Groove DNA Binding bZIP Proteins by Positive Patch Polyamides

The text of this chapter was taken in part from a publication coauthored with Ryan Bremer, Jason W. Szewczyk, and Prof. Peter Dervan.

*(R.E. Bremer, N.R. Wurtz, J.W. Szewczyk & P.B. Dervan *Bioorganic & Medicinal Chemistry.*, **2001**, 9, 2093-2103.)*

Abstract

Cell permeable synthetic ligands that bind to predetermined DNA sequences offer a chemical approach to gene regulation provided inhibition of a broad range of DNA transcription factors can be achieved. DNA minor groove binding polyamides containing aminoalkyl substituents at the N-1 of a single pyrrole residue display inhibitory effects for a bZIP proteins that bind exclusively in the DNA major groove. For major groove protein inhibition, specific protein-DNA contacts along the phosphate backbone were targeted with the positively charged dimethylamino substituent on the backbone of a minor groove binding hairpin. Remarkably, these polyamides bind DNA with enhanced affinity and uncompromised specificity when compared to polyamides with the aminoalkyl moiety at the C-terminus. By adding bZIP transcription factors to the class of protein-DNA complexes that can be disrupted by minor groove binding ligands, these results may increase the functional utility of polyamides modified at the N-alkyl position as regulators of gene expression.

Introduction

Hairpin polyamides containing pyrrole (Py), imidazole (Im) and hydroxypyrrole (Hp) amino acids are synthetic ligands that bind predetermined DNA sequences in the minor groove with affinities and specificities comparable to many DNA binding proteins (1, 2). Rules have been developed that allow for sequence specific recognition of the DNA minor groove by relating each Watson Crick base pair with a particular pairing of the aromatic Py, Im and Hp rings (1-4). The crescent shaped polyamides bind in the minor groove of DNA with pairs of aromatic rings stacked against each other and the walls of the groove, allowing the backbone amide hydrogens and the substituents at the 3-position of the Py, Im and Hp residues to make specific contacts with the edges of the intact base pairs. An γ -aminobutyric acid residue (γ) connects the polyamide subunits in a "hairpin" motif that enhances affinity and unambiguously locks the desired ring pairings in register.

A requirement for the general application of minor groove binding polyamides as regulators of gene transcription is motifs that can effectively inhibit all classes of DNA binding proteins, especially transcription factors which bind regulatory elements in gene promoter regions. Polyamides have been found to interfere with protein-DNA recognition in cases where contacts in the *minor* groove are important for protein-DNA binding affinity (2). Polyamides targeted to minor groove contacts of transcription factors Ets-1, LEF-1, and TBP inhibited DNA-binding of the each protein in mobility gel shift assays (5). In contrast, polyamides have been shown to bind simultaneously with some proteins such as bZIP protein, GCN4, that exclusively occupy the DNA *major* groove (6-8). The ubiquity of major groove contacts in protein:DNA recognition provides the impetus to

develop approaches for the inhibition of major groove proteins by DNA minor groove binding polyamides.

Hairpin polyamides with the tripeptide Arg-Pro-Arg at the C-terminus have been shown to inhibit GCN4, a bZIP protein bound exclusively in the major groove, potentially by competing with the protein side chains for electrostatic contacts to the DNA phosphate backbone (8). However, Arg-Pro-Arg modified polyamides may be limited by the use of α -amino acids with regard to biostability. This lead us to ask the question of whether a N-aminoalkylpyrrole residue could serve as a simple nonpeptide substitute for the Arg-Pro-Arg tripeptide positive patch that can be placed at any position along the polyamide backbone that would allow for discrete targeting of phosphodiester protein contacts.

Analysis of an X-ray crystal structure of a polyamide bound to its cognate DNA site reveals that the N-1 substituent of each pyrrole ring makes its closest phosphate contact two phosphates to the 3' side of the base the ring pair recognizes, consistent with the minor groove binding nature of the polyamide (9). Polyamides displaying an N-aminoalkyl positive patch could interfere with protein binding to this specific phosphate and consequently inhibit specific contacts with DNA. The increasing availability of protein-DNA crystal structures provide the necessary information to design a polyamide that targets important protein:DNA contacts. Previous attempts to inhibit major groove binding transcription factors with minor groove binding ligands have used polyamine substituents with ten charges attached to a tripyrrole, likely resulting in a ligand with significantly reduced sequence specificity (10).

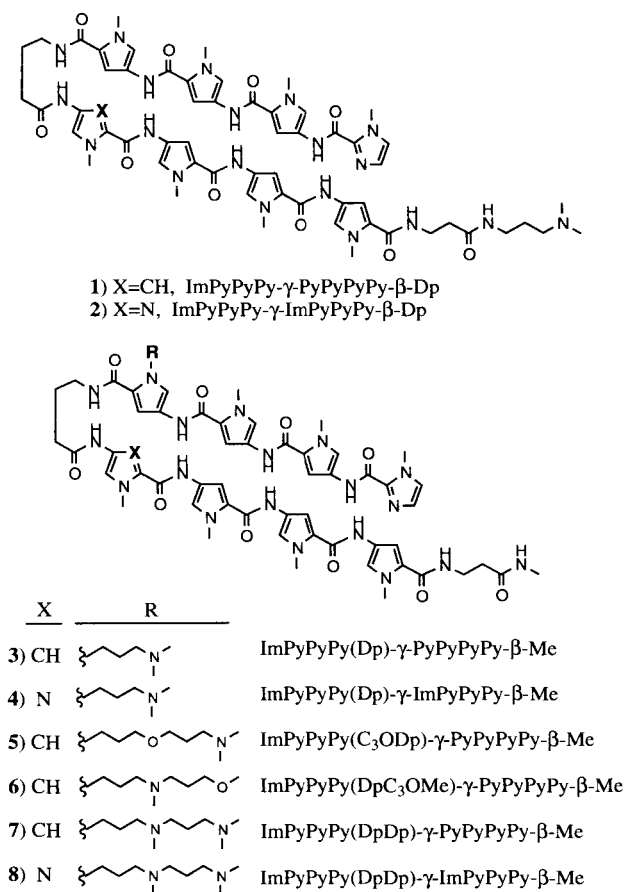


Figure 4.1. Structures of eight-ring hairpin polyamides containing N-methylpyrrole residues and a C-terminal Dp tail (top) and polyamides with N-aminoalkylpyrrole residues and a C-terminal N-methylamide. ImPyPyPy- γ -PyPyPyPy- β -Dp (1), and ImPyPyPy- γ -ImPyPyPy- β -Dp (2) (top). ImPyPyPy(Dp)- γ -PyPyPyPy- β -Me (3), ImPyPyPy(Dp)- γ -ImPyPyPy- β -Me (4), ImPyPyPy(C₃ODp)- γ -PyPyPyPy- β -Me (5), ImPyPyPy(DpC₃OMe)- γ -PyPyPyPy- β -Me (6), ImPyPyPy(DpDp)- γ -PyPyPyPy- β -Me (7) and ImPyPyPy(DpDp)- γ -ImPyPyPy- β -Me (8).

We report here the synthesis and DNA binding properties of a series of hairpin polyamides containing N-aminoalkyl substitutions at a single pyrrole residue and their ability to inhibit DNA binding by a major groove binding protein, GCN4 (222-281), a member of the bZIP family of transcriptional activators. Analysis of the crystal structure of GCN4-DNA complex reveals a key Lys 246-DNA phosphate backbone contact (11). Polyamides were designed to place a positive charge at the crucial DNA phosphate Lys246 position to diminish a likely important electrostatic interaction. A series of

polyamides were synthesized that contain multiple amines and selected ether substitutions as controls on the N-1 position of the pyrrole. The DNA binding affinity and specificity of each of these compounds for their target sites was evaluated using quantitative DNase I footprinting and compared to analogs containing the traditional N-methyl substituents. Gel mobility shift assays were employed to investigate the ability of these polyamides to inhibit DNA binding by a major groove binding protein.

Results

Polyamide Synthesis. The synthesis of N-aminoalkylpyrrole-containing polyamides **3-8** (Figure 1) required the preparation of a new pyrrole monomer suitable for solid phase synthesis and subsequent modification at the N-1 position. (3-Hydroxypropyl)-4-[(*tert*-butoxycarbonyl)amino]-pyrrole-2-carboxylic acid, (**11**), introduces a 3-hydroxypropyl moiety at the N-1 of pyrrole which can be modified by sulfonylation followed by nucleophilic displacement to afford a variety of derivatives from a single polyamide precursor. By cleaving the polyamide from resin with methylamine, polyamides (**3** and **4**) can be obtained that are structural isomers of those previously reported (**1** and **2**) and limit the overall charge of the N-diaminoalkylpyrrole-containing polyamides (**7** and **8**) (12). Ethyl 4-nitropyrrole-2-carboxylate (**9**) is available in 500 g quantities as described (13). Alkylation of the pyrrole N-1 with 3-iodopropanol provided **10**. Reduction and Boc protection of the amine followed by hydrolysis of the ethyl ester afforded **11** in gram quantities in 53% yield. (Figure 4.2)

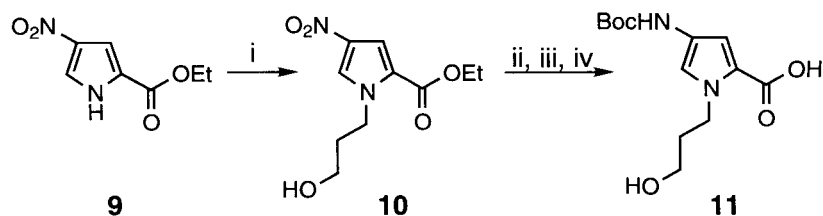


Figure 4.2. Synthesis of the Boc-*N*-(3-hydroxypropyl)pyrrole-acid monomer (**11**) for solid phase synthesis. (i) Iodopropanol, acetone, K_2CO_3 . (ii) H_2 , 10 % Pd/C, DMF. (iii) Boc-anhydride, DIEA. (iv) 1 M KOH, EtOH, 70 °C.

The *N*-(3-hydroxypropyl)pyrrole monomer **11** was used without further protection in the Boc-chemistry machine assisted solid phase synthesis of polyamides **12** and **13** from Boc- β -Ala-PAM resin (Figure 4.3) (14). In order to provide polyamides that differ from those previously reported only by the placement of the positively charged *N,N*-dimethylaminopropyl moiety and the *N*-methyl group, it was necessary to cleave the polyamide from resin using methylamine. This was accomplished by condensing methylamine in a Parr apparatus with resin-bound polyamide, followed by heating overnight (55 °C) to provide crude polyamides ImPyPyPy(C_3OH)- γ -PyPyPyPy- β -Me, (**12**), and ImPyPyPy(C_3OH)- γ -ImPyPyPy- β -Me, (**13**), containing the *N*-(3-hydroxypropyl)pyrrole residue suitable for further modification. The free hydroxyl of the crude polyamide was activated with *p*-toluenesulfonyl chloride or methanesulfonyl chloride in dry pyridine. Nucleophilic displacement of the tosylate or mesylate with the appropriate amine (**3**, **4**, **6-8**) or sodium salt (**5**) followed by preparatory reverse phase HPLC provided the desired *N*-aminoalkyl substituted polyamides in recoveries similar to those observed for *N*-methyl derivatives cleaved with dimethylaminopropylamine.

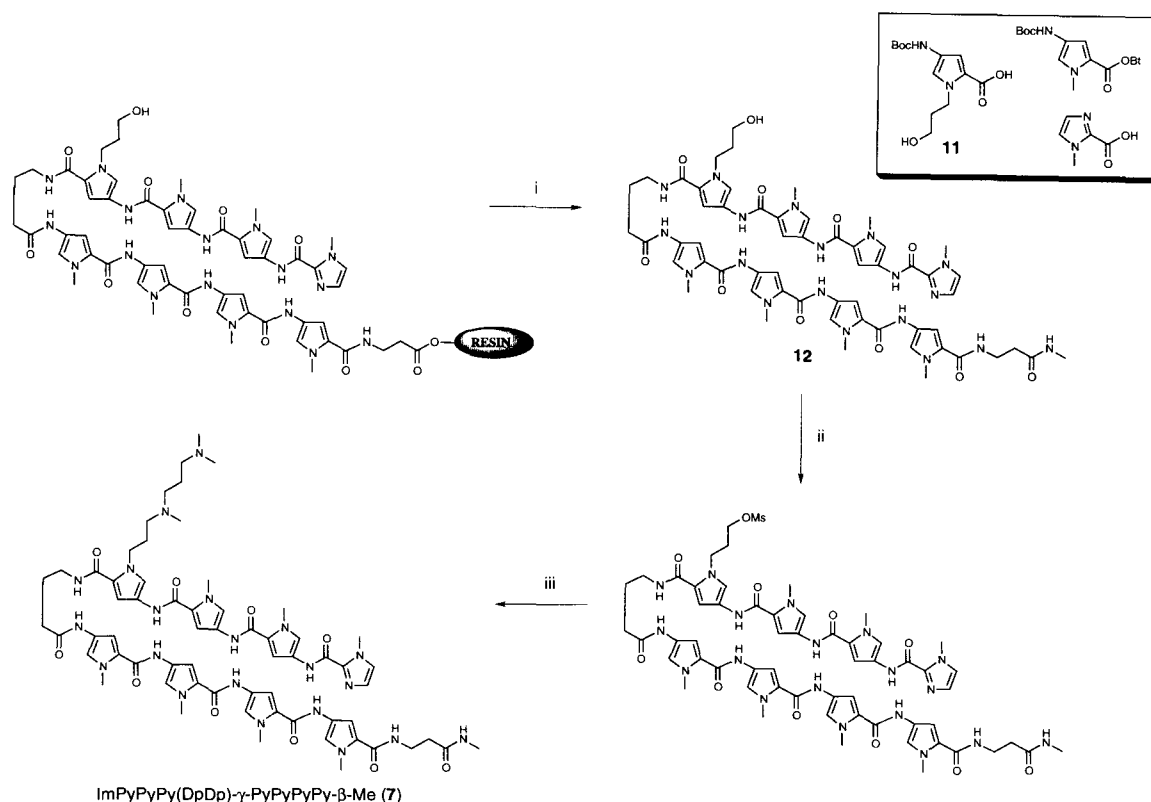


Figure 4.3. Synthetic scheme for solid phase preparation of *N*-aminoalkylpyrrole containing polyamides (15). Cycling protocols consisted of trifluoroacetic acid (TFA) deprotection followed by coupling with HOBt-activated Boc-pyrrole, Boc-imidazole, Boc-*N*-(3-hydroxypropyl)pyrrole (11) or aliphatic amino acid ester. (i) The PAM resin is then cleaved by treatment with methylamine. (ii) Activation with methanesulfonyl chloride or toluenesulfonyl chloride followed by reaction with the appropriate amine or alkoxide (iii) and reverse phase HPLC provided the desired products. The synthesis of ImPyPyPy(DpDp)- γ -PyPyPyPy- β -Me (7) is shown as a representative example.

Quantitative DNase I footprint titrations. The DNA binding affinity and specificity for the six base pair target sites were measured for the *N*-aminoalkylpyrrole-containing polyamides by quantitative DNase I footprinting on the 229-bp *AflIII/FspI* restriction fragment of pJT8 (12). Polyamides of the basic composition ImPyPyPy- γ -ImPyPyPy (2, 4 and 8) were designed to bind the site 5'-AGTACT-3' according to the pairing rules. Conversely, the single atomic substitution that converts an imidazole to a pyrrole residue provides polyamides based on ImPyPyPy- γ -PyPyPyPy (1, 3 and 5-7), targeted to 5'-AGTATT-3'. The sites 5'-AGTACT-3' and 5'-AGTATT-3' are for

ImPyPyPy- γ -ImPyPyPy polyamides “match” and “single base pair mismatch” sites, respectively, and for ImPyPyPy- γ -PyPyPyPy polyamides “single base pair mismatch” and “match” sites, respectively.

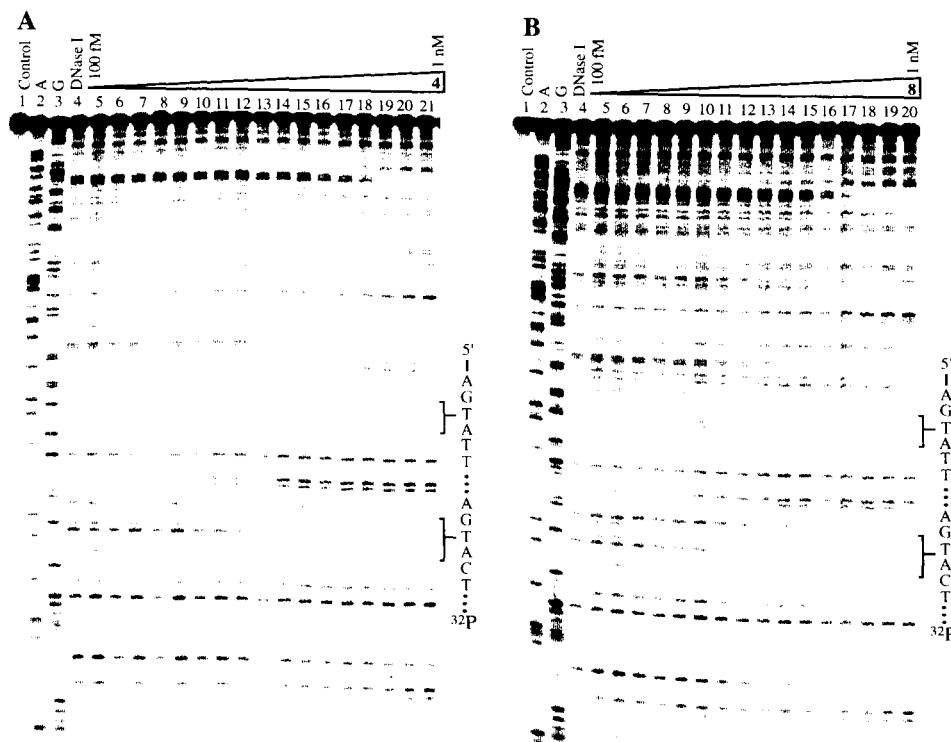


Figure 4.4. Storage phosphor autoradiograms of quantitative DNase I footprint titration experiments with (A) ImPyPyPy(Dp)- γ -ImPyPyPy- β -Dp (**4**) and (B) ImPyPyPy(DpDp)- γ -ImPyPyPy- β -Dp (**8**) on the 3'-³²P-end-labeled *AflIII/FspI* restriction fragment from pJT8 (*13*). All reactions contained 8 kcpm restriction fragment, 10 mM Tris•HCl (pH 7.0), 10 mM KCl, 10 mM MgCl₂ and 5 mM CaCl₂ and were performed at 22 °C. Lane 1, intact DNA; lane 2, A-specific reaction; lane 3, G-specific reaction; lane 4, DNase I standard; lanes 5-21, 100 fM, 200 fM, 500 fM, 1 pM, 1.5 pM, 2.5 pM, 4.0 pM, 6.5 pM, 10 pM, 15 pM, 25 pM, 40 pM, 65 pM, 100 pM, 200 pM, 500 pM, 1 nM, respectively of **3** or **8**. The 65 pM lane has been omitted for **8**. The positions of the match (5'-AGTACT-3') and single base pair mismatch (5'-AGTATT-3') sites are indicated.

DNase I footprinting revealed that ImPyPyPy(Dp)- γ -ImPyPyPy- β -Me (**4**) bound the 5'-AGTACT-3' with an equilibrium association constant $K_a = 1.9 (\pm 0.4) \times 10^{11} \text{ M}^{-1}$, and a 56-fold preference over the single base pair mismatch site ($K_a = 3.3 (\pm 0.4) \times 10^9 \text{ M}^{-1}$) (Figure 4.4A, Table 4.1). This represents ≈ 10 -fold increase in affinity relative to the parent polyamide differing in the placement of the dimethylaminopropyl and N-methyl

moieties, ImPyPyPy- γ -ImPyPyPy- β -Dp (**2**) ($K_a=1.4 (\pm 0.5) \times 10^{10} \text{ M}^{-1}$ for 5'-AGTACT-3'), and uncompromised specificity (61-fold, $K_a=2.3 (\pm 0.5) \times 10^8 \text{ M}^{-1}$ for 5'-AGTATT-3'). The increase in affinity upon placement of the alkylamine on the pyrrole ring was also observed for ImPyPyPy(Dp)- γ -PyPyPyPy- β -Me (**3**) relative to ImPyPyPy- γ -PyPyPyPy- β -Dp (**1**). Polyamide **3** bound the 5'-AGTATT-3' match site with an affinity $K_a = 1.3 (\pm 0.6) \times 10^{10} \text{ M}^{-1}$ and a 4-fold preference over the single base pair mismatch site ($K_a=2.9 (\pm 0.6) \times 10^9 \text{ M}^{-1}$). In comparison, the parent polyamide, **1**, bound both the match and mismatch sites with reduced affinity ($K_a=1.5 (\pm 0.3) \times 10^9 \text{ M}^{-1}$ for 5'-AGTATT-3' and $K_a=2.4 (\pm 1.2) \times 10^8 \text{ M}^{-1}$ for 5'-AGTACT-3').

In order to explore the potential applications of aminoalkyl substitutions of the pyrrole ring, the DNA binding affinities and specificities of analogs containing multiple amines or selected ether linkages were determined. The presence of an additional aminopropyl moiety to provide a polyamide with potentially two positive charges, as in ImPyPyPy(DpDp)- γ -ImPyPyPy- β -Me (**8**), resulted in an additional slight increase in affinity for the match site relative to **4** ($K_a=3.0 (\pm 0.4) \times 10^{11} \text{ M}^{-1}$), and a 36-fold preference over the mismatch site ($K_a=8.2 (\pm 3.1) \times 10^9 \text{ M}^{-1}$) (Figure 4.4B, Table 4.1). This represents only a modest decrease in specificity for **8** relative to **4** (36-fold and 56-fold, respectively). A similar decrease in specificity was observed for ImPyPyPy(DpDp)- γ -PyPyPyPy- β -Me (**7**) relative to polyamide **3** ($K_a=6.9 (\pm 1.2) \times 10^9 \text{ M}^{-1}$ for 5'-AGTATT-3' and $K_a=3.1 (\pm 1.0) \times 10^9 \text{ M}^{-1}$ for 5'-AGTACT-3'). Substitution of either tertiary amine in **7** to provide the singly charged polyamides ImPyPyPy(C₃ODp)- γ -PyPyPyPy- β -Me (**5**) and ImPyPyPy(DpC₃OMe)- γ -PyPyPyPy- β -Me (**6**) restored the optimal affinity and specificity observed with **3**, while retaining the overall size and length of **7**. Polyamide **5**,

containing an internal ether linkage and a terminal dimethylamine, bound the 5'-AGTATT-3' match site with a $K_a = 1.2 (\pm 0.1) \times 10^{10} \text{ M}^{-1}$ with a 7-fold preference over the 5'-AGTACT-3' mismatch ($1.8 (\pm 0.9) \times 10^9 \text{ M}^{-1}$). Similarly, the introduction of an internal amine and a terminal methyl ether in **6** resulted in $K_a = 1.5 (\pm 0.3) \times 10^{10} \text{ M}^{-1}$ and $2.2 (\pm 0.7) \times 10^9 \text{ M}^{-1}$ for the match and mismatch sites, respectively.

Table 4.1. Equilibrium association constants (M^{-1})^a

Polyamide	Match	Mismatch
ImPyPyPy- γ -PyPyPyPy- β -Dp (1)	$1.5 \times 10^9 (\pm 0.3)$	$2.4 \times 10^8 (\pm 1.2)$
ImPyPyPy(Dp)- γ -PyPyPyPy- β -Me (3)	$1.3 \times 10^{10} (\pm 0.6)$	$2.9 \times 10^9 (\pm 0.6)$
ImPyPyPy- γ -ImPyPyPy- β -Dp (2)	$1.4 \times 10^{10} (\pm 0.5)$	$2.3 \times 10^8 (\pm 0.5)$
ImPyPyPy(Dp)- γ -ImPyPyPy- β -Me (4)	$1.9 \times 10^{11} (\pm 0.4)$	$3.3 \times 10^9 (\pm 0.4)$
ImPyPyPy(C ₃ ODp)- γ -PyPyPyPy- β -Me (5)	$1.2 \times 10^{10} (\pm 0.1)$	$1.8 \times 10^9 (\pm 0.9)$
ImPyPyPy(DpC ₃ OMe)- γ -PyPyPyPy- β -Me (6)	$1.5 \times 10^{10} (\pm 0.3)$	$2.2 \times 10^9 (\pm 0.7)$
ImPyPyPy(DpDp)- γ -PyPyPyPy- β -Me (7)	$6.9 \times 10^9 (\pm 1.2)$	$3.1 \times 10^9 (\pm 1.0)$
ImPyPyPy(DpDp)- γ -ImPyPyPy- β -Me (8)	$3.0 \times 10^{11} (\pm 0.4)$	$8.2 \times 10^9 (\pm 0.4)$

^aValues reported are the mean values obtained from at least three DNase I footprint titration experiments. The assays were carried out at 22° C, 10 mM Tris•HCl (pH 7.0), 10 mM KCl, 10 mM MgCl₂, and 5 mM CaCl₂. The match and mismatch sites for **1**, **3** and **5-7** are 5'-ttAGTATTg-3' and 5'-ttAGTACTg-3', respectively. The match and mismatch sites for **2**, **4** and **8** are 5'-ttAGTACTg-3' and 5'-ttAGTATTg-3', respectively.

Protein Binding Inhibition by Polyamides determined by Gel Shift. The ability of N-aminoalkylpyrrole-containing polyamides to inhibit DNA binding by GCN4 (222-281) was investigated using gel mobility shift assays. The carboxy-terminal 60 amino acids of GCN4 (222-281) contain the “leucine zipper” dimerization domain and the basic region DNA binding domain and have been shown to be sufficient for sequence specific DNA binding (15, 16). GCN4 (222-281) binds its target site as a dimer of α -helices that make specific hydrogen bonds, van der Waal contacts and phosphate interactions in the DNA major groove (Figure 4A,B) (11, 17). Gel mobility shift assays were performed as previously described (6, 8, 16).

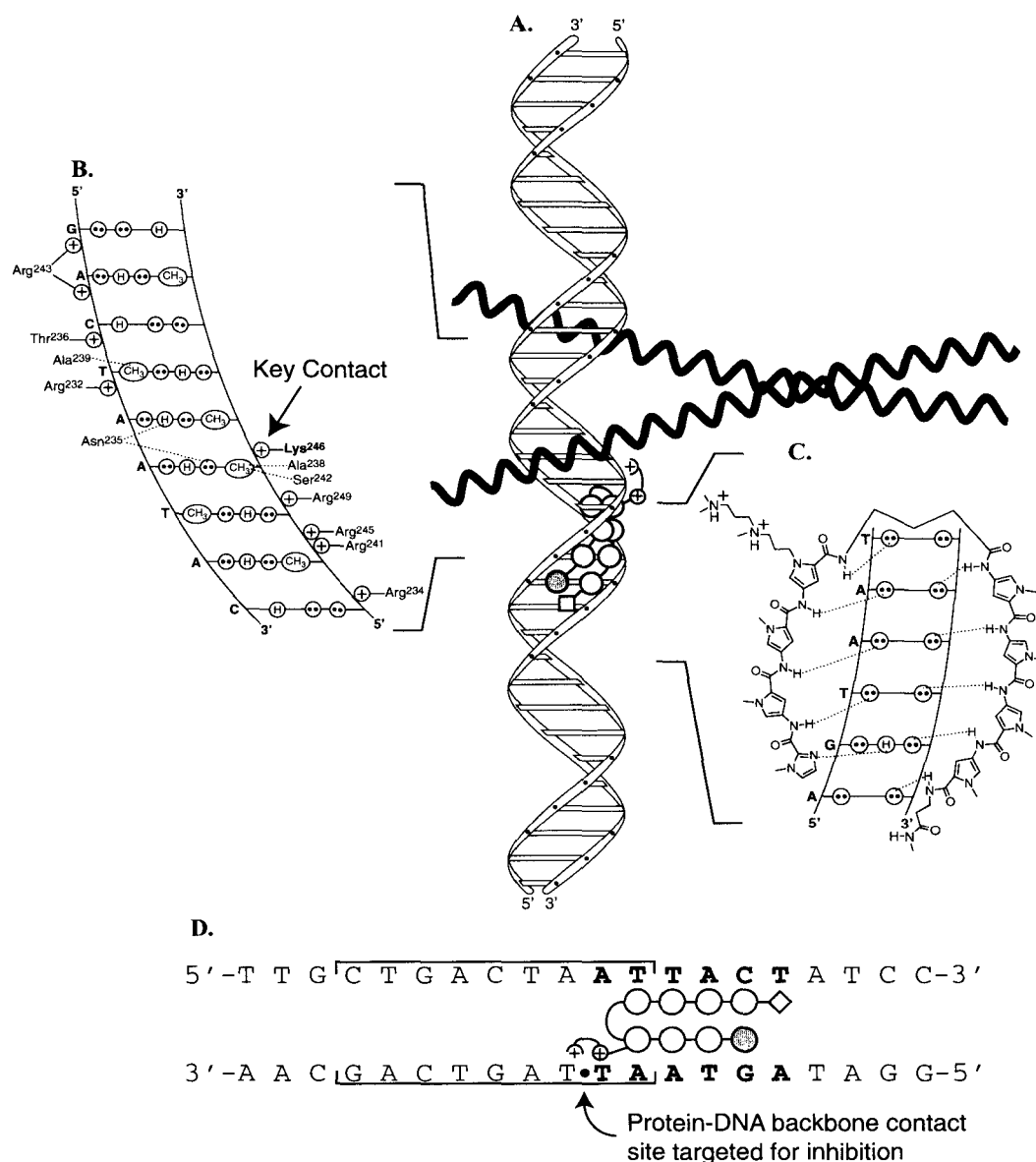


Figure 4.5. A schematic model of N-(aminoalkyl)pyrrole polyamides targeted to the major groove transcription factor, GCN4. (A) The α -helical GCN4 dimer is shown binding to adjacent major grooves (11). The N-(aminoalkyl)pyrrole polyamide is shown as black and white balls, which represent imidazole and pyrrole amino acids, respectively. The diamond represents β -alanine; γ -aminobutyric acid is designated as a curved line and the aminoalkyl chain is represented by a plus sign in a circle. (B) The contacts between one GCN4 monomer and the major groove of one half-site of 5'-CTGACTAAT-3' are depicted (adapted from Ref. #11). Circles with two dots represent the lone pairs of the N7 of purines, the O4 of thymine, and the O6 of guanine. Circles containing an H represent the N6 and N4 hydrogens of the exocyclic amines of adenine and cytosine, respectively. The C5 methyl group of thymine is depicted as a circle with CH₃ inside. Plus signs represent protein residues, which electrostatically contact the phosphate backbone. (C) The hydrogen bonding model of the eight-ring hairpin polyamide ImPyPyPy(DpDp)- γ -PyPyPyPy- β -Me bound to the minor groove of 5'-AGTAAT-3'. Circles with two dots represent the lone pairs of N3 of purines and O2 of pyrimidines. Circles containing an H represent the N2 hydrogens of guanines. Putative hydrogen bonds are illustrated by dotted lines. (D) The model of the polyamide binding its target site (bold) adjacent to the GCN4 binding site (brackets). Polyamide residues are as in A.

The polyamides of the sequence ImPyPyPy- γ -PyPyPyPy were designed to bind the 5'-AGTAAT-3' site adjacent to the GCN4 binding site (5'-CTGACTAAT-3') on the synthetic DNA duplex ARE-4. Aminoalkyl substitutions at the N-1 of the Py residue immediately to the N-terminal side of the γ -turn were designed to target the phosphate of the 3'-TpT-5' step (Figure 4.5). Incubation of the radiolabeled DNA fragment with the parent N-methyl containing polyamide ImPyPyPy- γ -PyPyPyPy- β -Dp (**1**) resulted in only a slight inhibition of protein binding, supporting our earlier observation that standard Py/Im polyamides bind simultaneously with major groove binding bZIP proteins (Figure 4.6A). The placement of *N*', *N*'-dimethylaminopropyl at the N-1 position of the specified pyrrole in **3** did not offer a significant improvement in GCN4 (222-281) inhibition over **1** (Figure 4.6B). However, incorporation of a second positive charge and a longer tether, as in **7**, provided 54% inhibition of GCN4 binding (Figure 4.6C, 4.7). The specificity of GCN4 inhibition was also tested using ImPyPyPy(DpDp)- γ -ImPyPyPy- β -Me (**8**), which is a single base pair mismatch for the polyamide site on ARE-4 and binds the DNA with significantly reduced affinity relative to **7**. Indeed, controls reveal that single mismatch polyamide **8** did not inhibit GCN4 binding (Figure 4.6D).

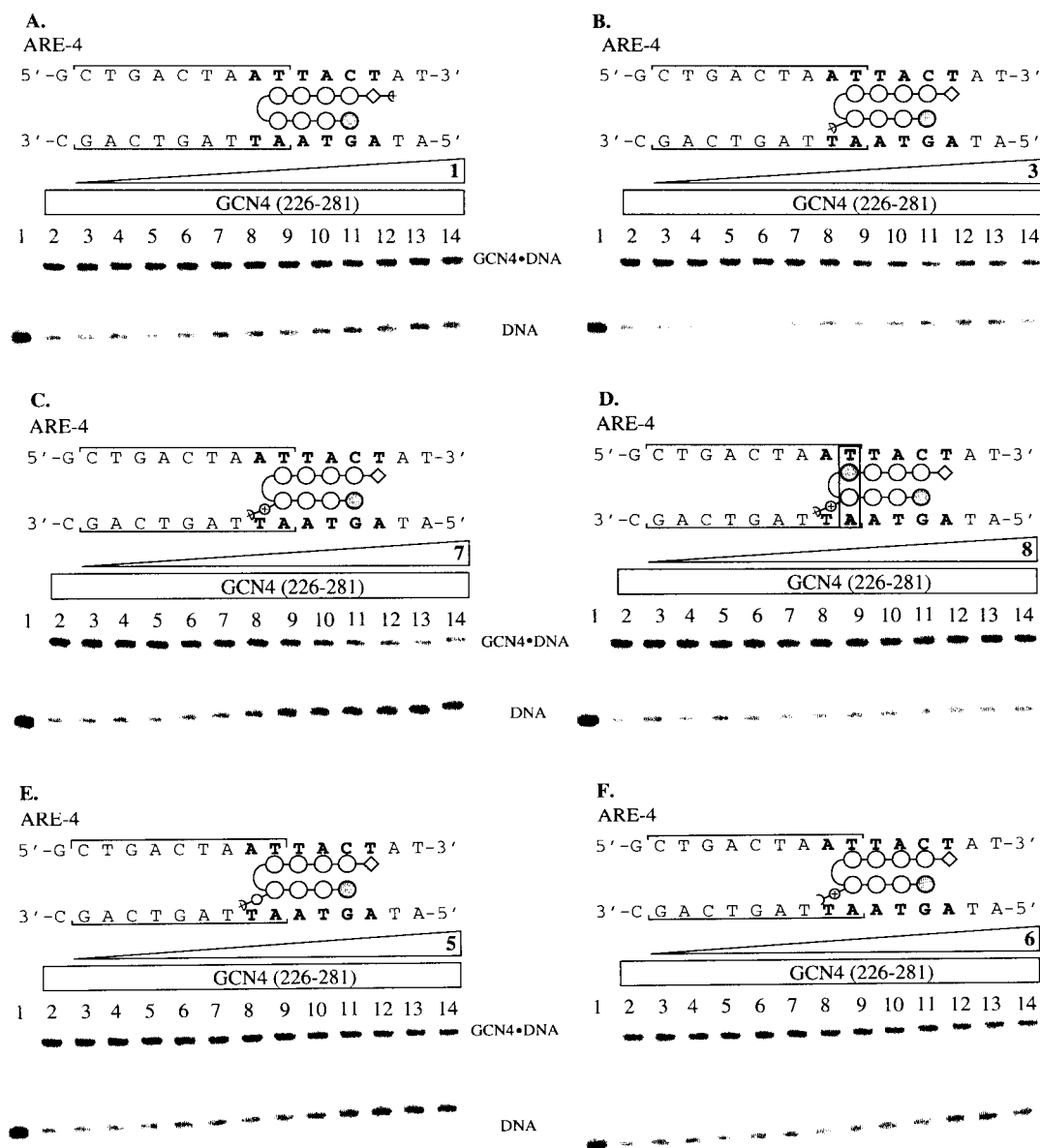


Figure 4.6. GCN4 (226-281) gel mobility shift experiments in the presence of N-methylpyrrole and N-aminoalkylpyrrole polyamides. (A-F) Top, model of polyamide binding the DNA fragment ARE-4. Mismatches are indicated by a grey box. Bottom, storage phosphor autoradiogram of non-denaturing polyacrylamide gel showing GCN4 (226-281) binding to the radiolabeled ARE-4 fragment in the presence of increasing concentrations of polyamide (10 mM bisTris pH 7.0, 100 mM NaCl, 1 mM DTT, 1 mM EDTA, 50 µg/mL poly(dI-dC)•poly(dI-dC), 22 °C). The upper band is the GCN4 (226-281)-DNA complex and the lower band is free DNA. Lane 1 is DNA only. Lane 2 contains DNA incubated with 200 nM GCN4 (226-281). Lanes 3-14 are 200 nM GCN4 (226-281) and 100 pM, 200 pM, 500 pM, 1 nM, 2 nM, 5 nM, 10 nM, 20 nM, 50 nM, 100 nM, 200 nM, and 500 nM polyamide: (A) ImPyPyPy-γ-PyPyPyPy-β-Dp (1), (B) ImPyPyPy(Dp)-γ-PyPyPyPy-β-Me (3), (C) ImPyPyPy(DpDp)-γ-PyPyPyPy-β-Me (7), (D) ImPyPyPy(DpDp)-γ-ImPyPyPy-β-Me (8), (E) ImPyPyPy(C₃ODp)-γ-ImPyPyPy-β-Me (5), (F) ImPyPyPy(DpC₃OMe)-γ-ImPyPyPy-β-Me (6).

Optimal GCN4 Inhibition Requires N-diaminoalkylpyrrole. It was important to determine if the basis for GCN4 inhibition by **7** was dependent upon the presence of either potentially charged amines or if perhaps the added length of the second aminopropyl was sufficient for inhibition. Polyamide **5** contains an ether linkage in place of the internal amine in **7**, resulting in a single charge placed distal to the polyamide. This change was sufficient to abolish GCN4 inhibition (Figure 4.6E). Furthermore, **6**, containing a single charge internal and a terminal methyl ether, exhibited similar behavior with only slight inhibition on GCN4 binding (Figure 4.6F). These results combined suggest that both of the positive charges present in **7** are necessary for inhibition of GCN4.

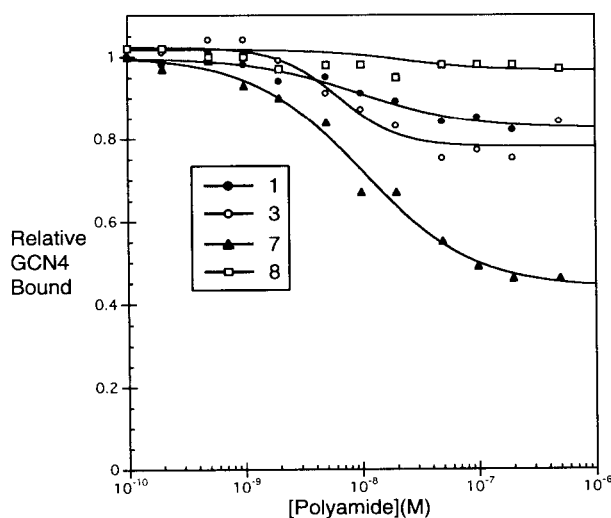


Figure 4.7. The level of GCN4 inhibition achieved by selected polyamides is compared. For each concentration of polyamide, the fraction of bound probe was determined by phosphorimager analysis followed by dividing the signal of the bound band by the sum of the bound and free bands. This value was normalized across experiments by dividing the value for each concentration by that for a control lane on the gel performed only with GCN4 (222-281). Typically, the fraction of bound probe in the absence of polyamide was ≈ 0.70 . The optimal GCN4 inhibition observed for **7** is depicted with triangles. The slight inhibition provided by **1** and **3** is shown as unfilled and filled circles, respectively. The absence of inhibition with **8** is depicted with squares.

Directing the Positive Patch to Adjacent Phosphates. To ask whether GCN4 binding inhibition by polyamide **7** was specific for a particular phosphate residue, DNA probes were designed to shift the polyamide binding site by one or two base pairs relative to that in ARE-4. It was necessary to not alter the DNA sequence of the protein binding site in this new design. To accomplish this, the A/T-tract at the 3' side of the GCN4 binding site and the degeneracy of Py/Py pairs for A•T and T•A base pairs was exploited. By simply shifting the G•C bp bound by the polyamide to the 3' or 5' side, probes were obtained that would deliver the polyamide positive patch to a different phosphate on the DNA backbone. ARE-3 and ARE-6 shift the polyamide to one or two base pairs deeper into the GCN4 binding site, respectively, while ARE-5 moves the polyamide one bp distal to the GCN4 cognate sequence, relative to ARE-4. Significant inhibition of GCN4 to any of these probes was not obtained at concentrations of **7** up to 20 times higher than that required for the maximum inhibition of GCN4 on ARE-4 (Figure 4.8). On ARE-3, **7** showed a slight amount of GCN4 inhibition (Figure 4.8C), while an apparent enhancement in GCN4 binding was observed on ARE-5 with as little as 100 nM **7** (Figure 4.8A). This study reveals that specific targeting of a protein-DNA contact is essential for successful inhibition of bZIP protein binding to DNA.

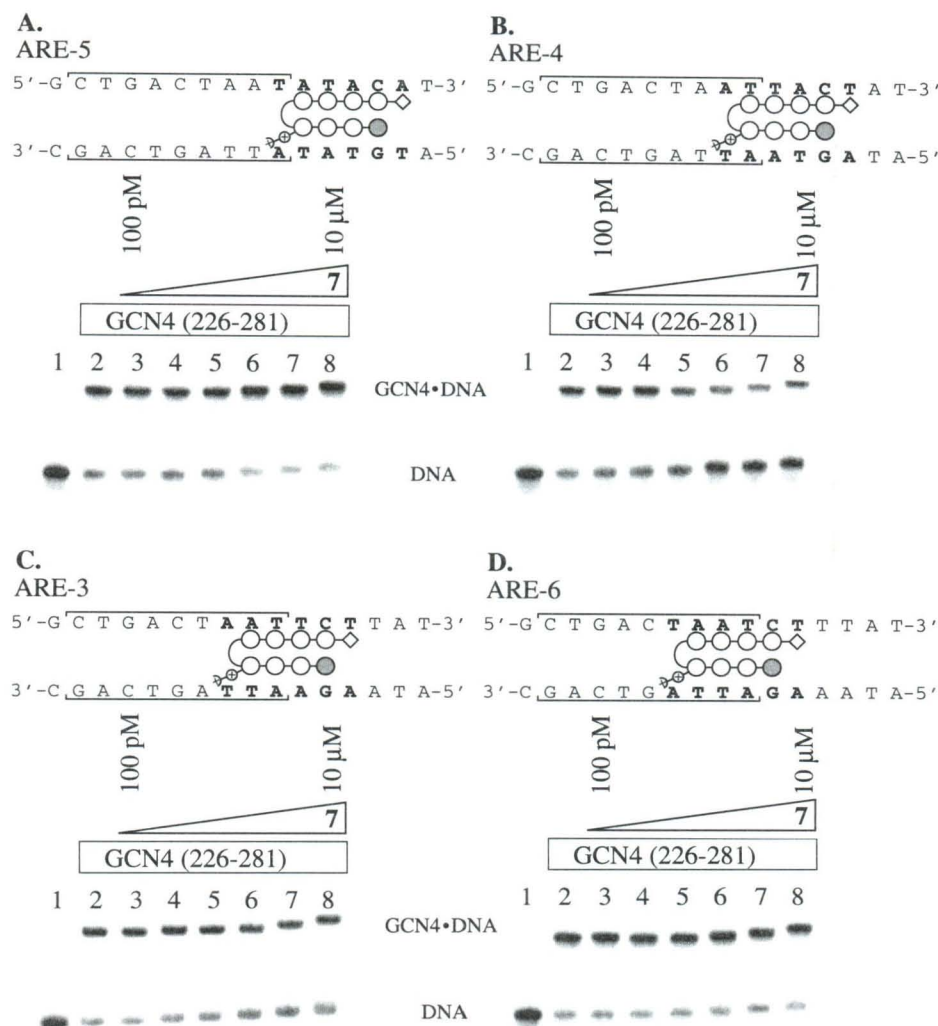


Figure 4.8. The effect of moving the polyamide binding site on GCN4 inhibition is investigated. Results are presented as in Figure 6. All panels include Lane 1: DNA only, Lane 2: 200 nM GCN4(226-281), Lane 3-8: 100 pM, 1 nM, 10 nM, 100 nM, 1 μM, and 10 μM **7**, respectively. (A) ARE-5 probe, (B) ARE-4 probe, (C) ARE-3 probe, (D) ARE-6 probe.

Discussion

A Py(Dp) Residue Increases Polyamide Affinity without Compromising Specificity. Generally, Py/Im polyamides are prepared using N-methyl aromatic amino acids and a dimethylaminopropylamide tail on the C-terminus as in **1** and **2**. Exchanging the placement of the Dp and N-Me to provide a polyamide with a *N*-(*N'*, *N'*-dimethylaminopropyl)pyrrole residue and a C-terminal N-methyl amide (**3** and **4**) affords

an isomer of identical molecular weight and composition; however, affinity for the DNA cognate site is increased 10-fold without compromising specificity at least in this case. The affinity enhancement observed may result from a specific interaction between the positively charged side chain and the phosphate oxygen or more likely relief of unfavorable steric interactions of the dimethylaminopropyl tail at the C-terminus clashing with the floor of the minor groove. As the generality of the pairing rules are explored for the sequence specific recognition of DNA, the incorporation of N-aminoalkylpyrrole residues to increase affinity may be useful for DNA sequences that otherwise cannot be targeted with subnanomolar affinity (18).

N-diaminoalkylpyrrole Polyamides Inhibit DNA Binding by GCN4. The incorporation of a doubly positively charged alkyl chain on the N-1 of a pyrrole residue has afforded a new class of hairpin polyamides that can effectively inhibit DNA binding by an exclusively major groove binding bZIP protein. The presence of the positive patch is designed to compete with protein side chains that make electrostatic contacts to the phosphate backbone, potentially by partially neutralizing the negative charge of the phosphate. The maximal GCN4 inhibition achieved with **7** is equivalent to that of the Arg-Pro-Arg analog; however, it is observed at 10-fold lower polyamide concentration for **7** than for the Arg-Pro-Arg polyamide (**8**). Furthermore, the N-diaminoalkyl moiety could be delivered from any ring in the polyamide, providing a versatility in design that allows for a polyamide target site optimized for recognition to be combined with the positive patch, offering a greater number of targetable major groove proteins.

The optimal location of the positive patch for GCN4 inhibition is consistent with X-ray crystal structure data and phosphate ethylation interference experiments (11, 19). As described above, the positive patch of the polyamide on ARE-4 is likely directed to the phosphate of the 3'-TpT-5' step. X-ray crystallography of GCN4 bound to a different DNA sequence reveals the residue Lys246 makes an electrostatic contact to the corresponding phosphate. Ethylation at the analogous phosphate on another GCN4 binding site resulted in a severe loss in affinity of GCN4 for the DNA target (19). While structural studies show other electrostatic contacts between protein side chains and the phosphates on either side of the 3'-TpT-5' step, ethylation of these phosphates has less of an effect on GCN4 binding affinity (11, 19). This latter observation is consistent with a lack of significant inhibition for **7** on ARE-3 and ARE-5. The slight inhibition observed for ARE-3 may be due to the flexible linker of **7** interacting with the 3'-TpT-5' step shown to inhibit GCN4 binding. Ethylation of the phosphate corresponding to the 3'-GpT-5' (ARE-6) reduced GCN4 affinity; however, no protein contacts to this phosphate have been identified by crystallography (11, 19). Taken together, these results suggest that the inhibition of a given protein:phosphate contact may not be universally equivalent in its effect on protein binding. Each contact plays a unique role in DNA recognition and inhibition of protein binding by positive patch polyamides may need to be optimized for each application.

The inability of polyamides containing ether linkages in place of the amines in **7** (**5** and **6**) to effectively inhibit GCN4 binding demonstrates at a minimum that an alkyl chain of this length is insufficient to obtain protein inhibition and that the amine nitrogens place a specific role in inhibiting GCN4 binding. The amines may be partially

neutralizing one or more phosphates to interfere with a protein:DNA contact or potentially distort the DNA by phosphate neutralization (20). Perhaps the strongest evidence for a specific phosphate contact comes from the protein inhibition data. It is not clear how the GCN4 inhibition by N-aminoalkylpyrrole polyamides could be explained in light of the previously discussed phosphate ethylation data in the absence of a specific polyamide:phosphate contact that partially neutralizes the negative charge of the phosphate. Modeling suggests that the diamino ligand does not possess the necessary length to significantly block the DNA major groove. If the protein were inhibited simply by steric bulk extending into the major groove, one would have expected **5** and/or **6** to effectively inhibit the protein. Furthermore, the lack of inhibition provided by the mismatched polyamide **8** demonstrates that the mechanism of inhibition requires a polyamide specifically bound in the minor groove.

Conclusions

A new class of hairpin polyamides containing a positively charged alkyl amine at the N-1 of a pyrrole ring have been shown to sequence specifically recognize DNA with high affinity and specificity relative to analogs with the amine at the C-terminus. The positive patch provided by a N-diaminoalkylpyrrole residue proved reasonably effective in the inhibition of DNA binding by the major groove binding protein GCN4 (222-281), potentially by competing with the protein side chains for electrostatic contacts to the DNA phosphate backbone. However, there remains room for improvement if we consider this series to be a first generation design. Nevertheless, these positive patch polyamides may expand the functional repertoire of polyamides for use in chemical genetics.

Experimental

All synthetic reagents were as previously described or obtained from Aldrich (14). Analytical HPLC was performed on a Beckman *Gold Nouveau* system with a model 126 pump and model 168 diode array detector. A Rainen C₁₈, Microsorb MV, 5 μ m, 300 \times 4.6 mm reverse phase column was employed with 0.1% (w/v) TFA:H₂O and 1.5% acetonitrile/min. Preparatory HPLC was performed on a Waters DeltaPak 25 \times 100 mm 100 μ m C₁₈ column in 0.1% (w/v) TFA, gradient elution 0.25%/min. CH₃CN. Resin substitution of synthesized polyamides was calculated as $L_{\text{new}}(\text{mmol/g}) = L_{\text{old}} / (1 + L_{\text{old}}(W_{\text{new}} - W_{\text{old}}) \times 10^{-3})$ where L is the loading (mmol of amine per gram of resin) and W is the weight (g mol⁻¹) of the growing peptide attached to the resin (14). DNA restriction fragment labeling, DNase I footprinting, and determination of equilibrium association constants were accomplished using previously described protocols (21, 22). Chemical sequencing reactions were performed according to published methods (23, 24).

Monomer Synthesis

Ethyl (3-hydroxypropyl)-4-nitropyrrole-2-carboxylate (10). Potassium carbonate (22.8 g, 163 mmol) was added to ethyl 4-nitropyrrole-2-carboxylate (**9**) (10 g, 54.3 mmol) in acetone (740 mL) and the slurry stirred for 2 hours, followed by the addition of 3-iodopropanol (23.5 g, 109 mmol) and refluxing for 2 hours. The reaction was cooled to room temperature, filtered and concentrated *in vacuo*. Water (200 mL) was added to the resulting oil, the pH of the aqueous layer was reduced to 3 with 10% H₂SO₄, and the mixture extracted with ethyl acetate. The combined ethyl acetate extracts were dried

(MgSO₄) and concentrated *in vacuo*. The solid was chromatographed with silica gel (ethyl acetate:hexanes) to provide a clear oil (12.8 g, 52.6 mmol, 99% yield): ¹H NMR (DMSO-*d*₆): δ 8.21 (s, 1H), 7.32 (s, 1H), 4.62 (t, 1H, J=4.9 Hz), 4.47 (t, 2H, J=6.8 Hz), 4.21 (q, 2H, J=7.0 Hz), 3.31 (q, 2H, J=6.1 Hz), 1.80 (m, 2H), 1.30 (t, 3H, J=7.0 Hz).

(3-Hydroxypropyl)-4-[(*tert*-butoxycarbonyl)amino]-pyrrole-2-carboxylic acid (11).

Ethyl (3-hydroxypropyl)-4-nitropyrrole-2-carboxylate (6.08 g, 25.1 mmol) was dissolved in DMF (20 mL). 10% Pd/C (1.0 g) was added and the mixture was stirred under hydrogen (300 psi) for 3 hours. Pd/C was removed by filtering through Celite. Di-*tert*-butyl dicarbonate (5.41 g, 24.8 mmol) and DIEA (8.6 mL, 50 mmol) was added and stirred for 2 hours. The reaction was cooled to room temperature, filtered, washed with methanol and concentrated *in vacuo*. Chromatography with silica gel (ethyl acetate:hexanes) provided a clear oil (5.10 g, 16.3 mmol, 65% yield): ¹H NMR (DMSO-*d*₆): δ 9.08 (s, 1H), 7.06 (s, 1H), 6.60 (s, 1H), 4.49 (t, 1H, J=5.1 Hz), 4.20 (t, 2H, J=6.8 Hz), 4.12 (q, 2H, J=7.5 Hz), 3.30 (q, 2H, J=5.5 Hz), 1.72 (m, 2H), 1.45 (s, 9H), 1.23 (t, 3H, J=7.5 Hz). Ethyl (3-hydroxypropyl)-4-[(*tert*-butoxycarbonyl)amino]-pyrrole-2-carboxylate (2.0 g, 6.4 mmol) was dissolved in ethanol (10 mL), followed by the addition of 1M KOH (40 mL) and heating (70 °C) for 2 hours. The reaction was cooled to room temperature and washed with ethyl ether. The pH of the aqueous layer was reduced to 3 with 10% H₂SO₄, and the mixture extracted with diethyl ether. The combined ethyl acetate extracts were dried (MgSO₄) and concentrated *in vacuo* to provide a white, crystalline solid (1.5 g, 5.3 mmol, 83% yield): ¹H NMR (DMSO-*d*₆): δ 12.04 (s, 1H), 9.05 (s, 1H), 7.04 (s, 1H), 6.54 (s, 1H), 4.20 (t, 2H, J=6.8 Hz), 3.27 (q, 2H, J=6.1 Hz), 1.73 (m, 2H), 1.40 (s, 9H).

Polyamide Synthesis

ImPyPyPy(C₃OH)- γ -PyPyPyPy- β -Me (12). ImPyPyPy(C₃OH)- γ -PyPyPyPy- β -PAM-resin was synthesized in a stepwise fashion by machine-assisted solid phase methods from Boc- β -PAM-resin (600 mg, 0.75 mmol/g). (3-hydroxypropyl)-4-[(*tert*-butoxycarbonyl)amino]-pyrrole-2-carboxylic acid (212 mg, 0.75 mmol), HBTU (284 mg, 0.75 mmol) and HOBt (304 mg, 2.25 mmol) were added dry to a synthesis cartridge. Upon delivery of DMF and DIEA activation occurs as described. A sample of polyamide resin (380 mg, 0.43 mmol/g) was cleaved with methylamine (55 °C, 28 hr.) in a Parr apparatus. The methylamine was allowed to evaporate and the crude polyamide was redissolved in 1:1 1M NH₄OH:CH₃CN (20 mL), filtered to remove the resin and lyophilized dry (146 μ mol).

ImPyPyPy(C₃OH)- γ -ImPyPyPy- β -Me (13). ImPyPyPy(C₃OH)- γ -ImPyPyPy- β -PAM-resin was synthesized in a stepwise fashion by machine-assisted solid phase methods from Boc- β -PAM-resin (450mg, 0.75 mmol/g) as described above. A sample of polyamide resin (178 mg, 0.43 mmol/g) was cleaved with methylamine (55 °C, 18 hr.) in a Parr apparatus. The methylamine was allowed to evaporate and the crude polyamide was redissolved in 1:1 1M NH₄OH:CH₃CN (20 mL), filtered to remove the resin and lyophilized dry (56 μ mol).

ImPyPyPy(Dp)- γ -ImPyPyPy- β -Me (3). Crude ImPyPyPy(C₃OH)- γ -ImPyPyPy- β -Me (25 μ mol) was dissolved in pyridine (600 μ L) followed by the addition of methanesulfonyl chloride (100 μ L). After 30 min., the reaction was precipitated with ethyl ether (1 mL), decanted, and redissolved in DMF (1 mL). 1:1 Dimethylamine (2.0M

in THF):DMF (3 mL) was added and the reaction allowed to proceed for 45 min. at 90 °C. The reaction was diluted to 8 mL with 0.1% (w/v) TFA and purified by preparatory reverse phase HPLC to afford ImPyPyPy(Dp)- γ -ImPyPyPy- β -Me (**3**) upon lyophilization of the appropriate fractions (1.3 mg, 1.0 μ mol, 4.0 % recovery); UV (H₂O) λ_{max} 312 (66,000); ¹H NMR (DMSO-*d*₆): δ 10.45 (s, 1H), 10.27 (s, 1H), 9.95 (s, 3H), 9.93 (s, 1H), 9.88 (s, 1H), 9.27 (m, 1H), 8.19 (m, 1H), 8.00 (m, 1H), 7.79 (m, 1H), 7.42 (s, 1H), 7.36 (s, 1H), 7.30 (s, 1H), 7.25 (s, 1H), 7.22 (s, 1H), 7.19 (s, 1H), 7.15 (s, 1H), 7.13 (s, 1H), 7.10 (s, 1H), 7.05 (s, 1H), 7.00 (s, 2H), 6.87 (s, 1H), 6.78 (s, 1H), 4.26 (m, 2H), 3.94 (s, 3H), 3.91 (s, 3H), 3.80 (s, 12H), 3.75 (s, 3H), 3.35 (m, 2H), 3.20 (m, 2H), 2.97 (m, 2H), 2.74 (d, 6H, J=4.8 Hz), 2.52 (d, 3H, J=4.5 Hz), 2.29 (m, 4H), 2.02 (m, 2H), 1.75 (m, 2H); MS (MALDI-TOF): [M+H]⁺ 1222.6, [M+Na]⁺ 1244.5, [M+K]⁺ 1260.5. (C₅₈H₇₂N₂₁O₁₀⁺, [M+H]⁺, calc. 1222.6; C₅₈H₇₁N₂₁NaO₁₀⁺, [M+Na]⁺, calc. 1244.6; C₅₈H₇₁KN₂₁O₁₀⁺, [M+K]⁺, calc. 1260.5).

ImPyPyPy(Dp)- γ -PyPyPyPy- β -Me (4**).** Crude ImPyPyPy(C₃OH)- γ -PyPyPyPy- β -Me (49 μ mol) was dissolved in pyridine followed by the addition of toluenesulfonyl chloride (2.2 M in pyridine, 350 μ L) and kept at 0 °C for 10 hours. After the addition of dimethylamine (2.0 M in THF, 2.0 mL), the reaction was heated at increasing temperatures until it was determined complete by analytical HPLC after 5 hours at 80 °C. The reaction mixture was diluted to 8 mL with 0.1% (w/v) TFA and purified by preparatory reverse phase HPLC to provide ImPyPyPy(Dp)- γ -PyPyPyPy- β -Me (**4**) upon lyophilization of the appropriate fractions (5.8 mg, 4.8 μ mol, 9.8% recovery); UV (H₂O) λ_{max} 312 (66,000); ¹H NMR (DMSO-*d*₆): δ 10.46 (s, 1H), 9.96 (s, 1H), 9.95 (s, 1H), 9.92 (s, 1H), 9.89 (s, 1H), 9.88 (s, 1H), 9.83 (s, 1H), 9.2 (br s, 1H), 8.18 (m, 1H), 7.99 (m,

1H), 7.80 (m, 1H), 7.38 (s, 1H), 7.31 (d, 1H, $J=1.8$ Hz), 7.26 (d, 1H, $J=1.8$ Hz), 7.20 (d, 1H, $J=1.5$ Hz), 7.16 (m, 3H), 7.06 (d, 1H, $J=1.8$ Hz), 7.03 (s, 1H), 7.02 (s, 1H), 7.01 (s, 1H), 6.89 (d, 1H, $J=1.5$ Hz), 6.84 (d, 1H, $J=1.5$ Hz), 6.79 (d, 1H, $J=1.5$ Hz), 4.24 (m, 2H), 3.96 (s, 3H), 3.82 (s, 9H), 3.81 (s, 3H), 3.80 (s, 3H), 3.76 (s, 3H), 3.35 (m, 2H), 3.20 (m, 2H), 3.00 (m, 2H), 2.74 (d, 6H, $J=5.1$ Hz), 2.53 (d, 3H, $J=4.5$ Hz), 2.27 (m, 2H), 2.03 (m, 2H), 1.97 (m, 2H), 1.78 (m, 2H); MS (MALDI-TOF): $[M+H]^+$ 1221.6. ($C_{59}H_{73}N_{20}O_{10}^+$, $[M+H]^+$, calc. 1221.6).

ImPyPyPy(C₃ODp)- γ -PyPyPyPy- β -Me (5). Crude ImPyPyPy(C₃OH)- γ -PyPyPyPy- β -Me (24 μ mol) in pyridine (0.4 mL) at 0 °C was activated with MsCl (75 μ L) at 0° C for three hours, followed by precipitation with ether (1 mL). The supernatant was redissolved in DMF (0.5 mL) and cooled to 0° C. *N,N*-dimethyl-3-amino-propanol (14 μ L, 120 μ mol) was added to NaH (60% dispersion in mineral oil, 4.7 mg, 114 μ mol) in DMF (150 μ L) and added to the reaction. Repeating the addition of NaH and *N,N*-dimethyl-3-amino-propanol and warming to room temperature did not provide sufficient reaction after 27 hours as determined by analytical HPLC. More NaH (60% dispersion in mineral oil, 60 mg, 1.5 mmol) and *N,N*-dimethyl-3-amino-propanol (180 μ L, 1.5 mmol) was added, followed by reaction at room temperature for 22 hours. The reaction mixture was diluted to 8 mL with 0.1% (w/v) TFA and purified by preparatory reverse phase HPLC to afford ImPyPyPy(C₃ODp)- γ -PyPyPyPy- β -Me (5) upon lyophilization of the appropriate fractions (1.4 mg, 1.1 μ mol, 4.6 % recovery); UV (H₂O) λ_{max} 312 (66,000); ¹H NMR (DMSO-*d*₆): δ 10.43 (s, 1H), 9.94 (s, 1H), 9.93 (s, 1H), 9.90 (s, 1H), 9.87 (s, 2H), 9.81 (s, 1H), 9.25 (m, 1H), 8.16 (m, 1H), 7.97 (m, 1H), 7.77 (m, 1H), 7.35 (s, 1H), 7.32 (s, 1H), 7.30 (s, 1H), 7.24 (s, 1H), 7.18 (s, 3H), 7.14 (s, 1H), 7.13 (s, 2H), 7.05 (s, 1H), 7.00 (s,

2H), 6.88 (s, 1H), 6.82 (s, 1H), 6.77 (s, 1H), 4.25 (m, 2H), 3.94 (s, 3H), 3.80 (s, 15H), 3.75 (s, 3H), 3.34 (m, 4H), 3.20 (m, 2H), 3.00 (m, 2H), 2.72 (d, 6H, $J=4.8$ Hz), 2.52 (d, 3H, $J=4.2$ Hz), 2.24 (m, 6H), 2.02 (m, 2H), 1.76 (m, 4H); MS (MALDI-TOF): $[M+H]^+$ 1279.7. ($C_{62}H_{79}N_{20}O_{11}^+$, $[M+H]^+$, calc. 1279.6).

ImPyPyPy(DpC₃OMe)- γ -PyPyPyPy- β -Me (6). MsCl activation of crude ImPyPyPy(C₃OH)- γ -PyPyPyPy- β -Me (24 μ mol) as for **4** followed by redissolving of the precipitate in DMF (3 mL), reaction with 3-methoxypropylamine (1 mL, 90° C, 30 min.) and preparatory reverse phase HPLC afforded ImPyPyPy(DpC₃OMe)- γ -PyPyPyPy- β -Me (**6**) upon lyophilization of the appropriate fractions (1.5 mg, 1.2 μ mol, 5.0% recovery); UV (H₂O) λ_{max} 312 (66,000); ¹H NMR (DMSO-*d*₆): δ 10.44 (s, 1H), 9.96 (s, 1H), 9.95 (s, 1H), 9.91 (s, 1H), 9.88 (s, 1H), 9.88 (s, 1H), 9.83 (s, 1H), 8.24 (m, 2H), 8.20 (m, 1H), 7.99 (m, 1H), 7.80 (m, 1H), 7.37 (d, 1H, $J=0.6$ Hz), 7.32 (d, 1H, $J=1.8$ Hz), 7.26 (d, 1H, $J=1.8$ Hz), 7.20 (m, 3H), 7.16 (m, 3H), 7.06 (d, 1H, $J=1.8$ Hz), 7.01 (m, 3H), 6.89 (d, 1H, $J=1.5$ Hz), 6.84 (d, 1H, $J=1.8$ Hz), 6.79 (d, 1H, $J=2.1$ Hz), 4.30 (m, 2H), 3.96 (s, 3H), 3.82 (s, 9H), 3.81 (s, 3H), 3.81 (s, 3H), 3.76 (s, 3H), 3.19 (s, 3H), 2.90 (m, 4H), 2.52 (m, 5H), 2.42 (m, 2H), 2.27 (m, 6H), 2.00 (m, 2H), 1.78 (m, 4H); MS (MALDI-TOF): $[M+H]^+$ 1265.8, $[M+Na]^+$ 1287.8, $[M+K]^+$ 1303.8. ($C_{61}H_{76}N_{20}O_{11}^+$, $[M+H]^+$, calc. 1265.6; $C_{61}H_{76}N_{20}NaO_{11}^+$, $[M+Na]^+$, calc. 1287.6; $C_{61}H_{76}KN_{20}O_{11}^+$, $[M+K]^+$, calc. 1303.6).

ImPyPyPy(DpDp)- γ -PyPyPyPy- β -Me (7). MsCl activation of crude ImPyPyPy(C₃OH)- γ -PyPyPyPy- β -Me (23 μ mol) as for **4** followed by redissolving of the precipitate in pyridine (0.75 mL), reaction with *N,N,N'*-trimethyl-1,3-propanediamine (0.75 mL, 90 °C, 40 min.) and preparatory reverse phase HPLC afforded ImPyPyPy(DpDp)- γ -PyPyPyPy- β -Me (**7**) upon lyophilization of the appropriate fractions (6.7 mg, 5.2 μ mol, 23%

recovery); UV (H₂O) λ_{max} 312 (66,000); ¹H NMR (DMSO-*d*₆): δ 10.45 (s, 1H), 9.97 (s, 1H), 9.95 (s, 1H), 9.92 (s, 1H), 9.89 (s, 1H), 9.88 (s, 1H), 9.84 (s, 1H), 9.55 (m, 1H), 8.44 (m, 1H), 8.19 (m, 1H), 7.98 (m, 1H), 7.80 (m, 1H), 7.37 (m, 1H), 7.34 (d, 1H, *J*=1.8 Hz), 7.26 (d, 1H, *J*=1.8 Hz), 7.20 (s, 3H), 7.15 (m, 2H), 7.07 (d, 1H, *J*=1.5 Hz), 7.01 (m, 3H), 6.88 (d, 1H, *J*=1.8 Hz), 6.84 (d, 1H, *J*=1.8 Hz), 6.79 (d, 1H, *J*=1.8 Hz), 4.30 (m, 2H), 3.96 (s, 3H), 3.82 (s, 9H), 3.81 (s, 6H), 3.80 (s, 3H), 3.35 (m, 4H), 3.20 (m, 2H), 3.00 (m, 2H), 2.75 (m, 9H), 2.54 (d, 3H, *J*=4.5 Hz), 2.26 (m, 6H), 2.04 (m, 2H), 1.96 (m, 2H), 1.77 (m, 2H); MS (MALDI-TOF): [M+H]⁺ 1292.7 (C₆₃H₈₂N₂₁O₁₀⁺, [M+H]⁺, calc. 1292.7).

ImPyPyPy(DpDp)- γ -ImPyPyPy- β -Me (8). Crude ImPyPyPy(C₃OH)- γ -ImPyPyPy- β -Me (23 μ mol) was dissolved in pyridine (150 μ L) and activated with TsCl (4.8M in pyridine, 200 μ L) for 1 hour at 0 °C. The addition of *N,N,N'*-trimethyl-1,3-propanediamine (0.30 mL, 55 °C, 30 min.) followed by preparatory reverse phase HPLC afforded ImPyPyPy(DpDp)- γ -ImPyPyPy- β -Me (**8**) upon lyophilization of the appropriate fractions (0.4 mg, 0.3 μ mol, 1.3% recovery); UV (H₂O) λ_{max} 312 (66,000); ¹H NMR (DMSO-*d*₆): δ 10.43 (s, 1H), 10.26 (s, 1H), 9.96 (s, 1H), 9.93 (s, 1H), 9.91 (s, 2H), 9.86 (s, 1H), 9.45 (m, 1H), 8.17 (m, 1H), 7.97 (m, 1H), 7.78 (m, 1H), 7.41 (s, 1H), 7.36 (s, 2H), 7.33 (s, 1H), 7.25 (s, 1H), 7.22 (s, 1H), 7.18 (s, 2H), 7.14 (s, 1H), 7.14 (s, 1H), 7.11 (s, 1H), 7.00 (s, 2H), 6.86 (s, 1H), 6.77 (s, 1H), 4.24 (m, 2H), 3.96 (s, 3H), 3.91 (s, 3H), 3.80 (s, 12H), 3.75 (s, 3H), 3.35 (m, 4H), 3.20 (m, 2H), 3.00 (m, 2H), 2.74 (d, 9H, *J*=3.6 Hz), 2.52 (d, 3H, *J*=4.8 Hz), 2.27 (m, 6H), 2.05 (m, 2H), 1.94 (m, 2H), 1.78 (m, 2H); MS (MALDI-TOF): [M+H]⁺ 1293.9 (C₆₂H₈₁N₂₂O₁₀⁺, [M+H]⁺, calc. 1293.7).

Acknowledgments

We are grateful to the National Institutes of Health for research support, the National Science Foundation for a predoctoral fellowship to R.E.B., Bristol-Myers Squibb and the Ralph M. Parson Foundation for predoctoral fellowships to R.E.B. and N.R.W., and the National Institutes of Health for a research service award to J.W.S. We thank G.M. Hathaway and the Caltech Protein/Peptide Microanalytical Laboratory for MALDI-TOF mass spectrometry.

References

1. Dervan, P. B., and Bürli, R. W. (1999) *Curr. Opin. Chem. Biol.* 3, 688-693.
2. Gottesfeld, J. M., Turner, J. M., and Dervan, P. B. (2000) *Gene Expression* 9, 77-91.
3. White, S., Baird, E. E., and Dervan, P. B. (1997) *Chem. Biol.* 4, 569-578.
4. White, S., Szewczyk, J. W., Turner, J. M., Baird, E. E., and Dervan, P. B. (1998) *Nature* 391, 468-471.
5. Dickinson, L. A., Gulizia, R. J., Trauger, J. W., Baird, E. E., Mosier, D. E., Gottesfeld, J. M., and Dervan, P. B. (1998) *Proc. Natl. Acad. Sci. USA* 95, 12890-12895.
6. Oakley, M. G., Mrksich, M., and Dervan, P. B. (1992) *Biochemistry* 31, 10969-10975.
7. Neely, L., Trauger, J. W., Baird, E. E., Dervan, P. B., and Gottesfeld, J. M. (1997) *J. Mol. Biol.* 274, 439-445.
8. Bremer, R. E., Baird, E. E., and Dervan, P. B. (1998) *Chem. Biol.* 5, 119-133.
9. Kielkopf, C. L., Baird, E. E., Dervan, P. D., and Rees, D. C. (1998) *Nat. Struct. Biol.* 5, 104-109.
10. Bruice, T. C., Sengupta, D., Blasko, A., Chiang, S. Y., and Beerman, T. A. (1997) *Bioorganic & Medicinal Chemistry* 5, 685-692.
11. Ellenberger, T. E., Brandl, C. J., Struhl, K., and Harrison, S. C. (1992) *Cell* 71, 1223-1237.
12. Trauger, J. W., Baird, E. E., and Dervan, P. B. (1996) *Nature* 382, 559-561.

13. Bremer, R. E., Szewczyk, J. W., Baird, E. E., and Dervan, P. B. (2000) *Bioorganic & Medicinal Chemistry* 8, 1947-1955.
14. Baird, E. E., and Dervan, P. B. (1996) *J. Am. Chem. Soc.* 118, 6141-6146.
15. Hope, I. A., and Struhl, K. (1986) *Cell* 46, 885-894.
16. Oakley, M. G., and Dervan, P. B. (1990) *Science* 248, 847-850.
17. Konig, P., and Richmond, T. J. (1993) *J. Mol. Biol.* 233, 139-154.
18. Swalley, S. E., Baird, E. E., and Dervan, P. B. (1997) *J. Am. Chem. Soc.* 119, 6953-6961.
19. Gartenberg, M. R., Ampe, C., Steitz, T. A., and Crothers, D. M. (1990) *Proc. Natl. Acad. Sci. USA* 87, 6034-6038.
20. Maher III, L. J. (1998) *Curr. Opin. Chem. Biol.* 2, 688-694.
21. Turner, J. M., Baird, E. E., and Dervan, P. B. (1997) *J. Am. Chem. Soc.* 119, 7636-7644.
22. Swalley, S. E., Baird, E. E., and Dervan, P. B. (1997) *Chem.-Eur. J.* 3, 1600-1607.
23. Maxam, A. M., and Gilbert, W. S. (1980) *Methods Enzymol.* 65, 499-560.
24. Iverson, B. L., and Dervan, P. B. (1987) *Nucleic Acids Res.* 15, 7823-7830.

CHAPTER 5

Inhibition of DNA Binding by NF- κ B with Pyrrole-Imidazole

Hairpin Polyamides

Abstract

Cell permeable synthetic ligands that bind to predetermined DNA sequences will offer a chemical approach to gene regulation if inhibition of a broad range of transcription factors can be achieved. Pyrrole/imidazole hairpin polyamides inhibit DNA binding by NF- κ B at nanomolar concentrations. Inhibition was observed only for the ligands that bind the 5'-GGGAC-3' subsite and not the 5'-TTCC-3' subsite of the heterodimeric protein binding site. Polyamides that bind the NF- κ B site incorporate new motifs for high binding of G,C rich sequences that have not been achieved using previously reported molecules. Addition of NF- κ B to the list of protein-DNA complexes that can be disrupted by minor groove binding ligands potentially increases the utility of polyamides as regulators of gene expression.

Introduction

Hairpin polyamides containing pyrrole (Py), imidazole (Im) and hydroxypyrrole (Hp) amino acids are synthetic ligands that bind predetermined DNA sequences in the minor groove with affinities and specificities comparable to many DNA binding proteins (1, 2). Rules have been developed that allow for sequence specific recognition of the DNA minor groove by relating each Watson Crick base pair with a particular pairing of the aromatic Py, Im and Hp rings (1-4). The crescent shaped polyamides bind in the minor groove of DNA with pairs of aromatic rings stacked against each other and the walls of the groove, allowing the backbone amide hydrogens and the substituents at the 3-position of the Py, Im and Hp residues to make specific contacts with the edges of the intact base pairs. An γ -aminobutyric acid residue (γ) connects the polyamide subunits in a "hairpin" motif that enhances affinity and unambiguously locks the desired ring pairings in register (1, 2).

In order for minor groove binding polyamides to be utilized as regulators of gene transcription, strategies must be developed that permit inhibition of all classes of DNA binding proteins, particularly transcription factors which bind regulatory elements in gene promoter regions. Working toward this goal, polyamides have been found to interfere with DNA binding for several classes of proteins (5). An examination of the available crystal structures of these proteins reveals that inhibition usually occurs when a hairpin polyamide targets a key protein-DNA contact in the minor groove. In an initial study a polyamide inhibited TFIIIA binding when it was bound to the position where one of TFIIIA's zinc fingers crosses the minor groove (6, 7). In other cases polyamides inhibited targeted TATA Box-binding protein, TBP, and LEF-1, when bound immediately adjacent

to their binding sites. Both of these proteins bind DNA in the minor groove and bend and unwind the DNA helix (8). It is believed that polyamides lock the DNA into a B-type formation and serve as a steric blockage, preventing protein binding. In the case of winged-helix-turn-helix transcription factor ETS-1, binding by the protein was inhibited when a polyamide was targeted to a loop that binds across the minor groove (8, 9). In contrast, polyamides have been shown to bind simultaneously with some proteins such as bZIP protein, GCN4, that exclusively occupy the DNA major groove, unless key phosphate contacts are targeted with a ligand with multiple positive charges displayed on a polyamide (10, 11). In this case we learned that it was necessary to target a specific phosphate for optimal inhibition. The DNA binding of other proteins has also been successfully inhibited, but many other protein classes have yet to be targeted by polyamides (see Table 5.1). Each new class of proteins presents a different challenge for designing minor-groove binding polyamides that can inhibit protein-DNA binding, especially for proteins that predominantly dock to the major groove of DNA.

Table 5.1. Protein Targets for Polyamide Inhibitors

Protein	DNA-Binding Motif
TFIIIA	zinc finger
TBP	minor groove/saddle
LEF-1	HMG Box
Ets-1	winged-helix-turn-helix
IE-86	minor groove
Estrogen Receptor	C4 zinc finger
Deadpan	basic helix-loop-helix
GCN4	bZip
Tax	minor groove

NF- κ B p50/p65 is a heterodimeric protein that was first characterized as a nuclear factor that binds to the κ light-chain enhancer in B-lymphocytes (12). Subsequently, NF- κ B has been found to activate transcription as part of signal transduction pathways in many different cell types and regulates genes involved in immune and inflammatory cell function (13, 14) and in anti-apoptotic responses (15-18). While p50 and p65 can bind as both homo- and heterodimers, the p50/p65 heterodimer is most prevalent (19). The DNA-binding domains of both p50 and p65 are located near their N-termini and are members of the Rel homology domain family of DNA-binding domains, consisting of two immunoglobulin-like β -sandwich domains joined by a linker region with the C-terminal domain mediating dimerization contacts (20, 21). DNA contacts are formed with loops joining β -strands in both the N- and C-terminal domains. The heterodimer binds 5'-GGGACTTCC-3'. The p65 monomer derives its specificity from contacts in the major groove for the four-base pair subsite (5'-TTCC-3'), while the p50 monomer recognizes a five base-pair subsite (5'-GGGAC-3'). The heterodimer makes several phosphate

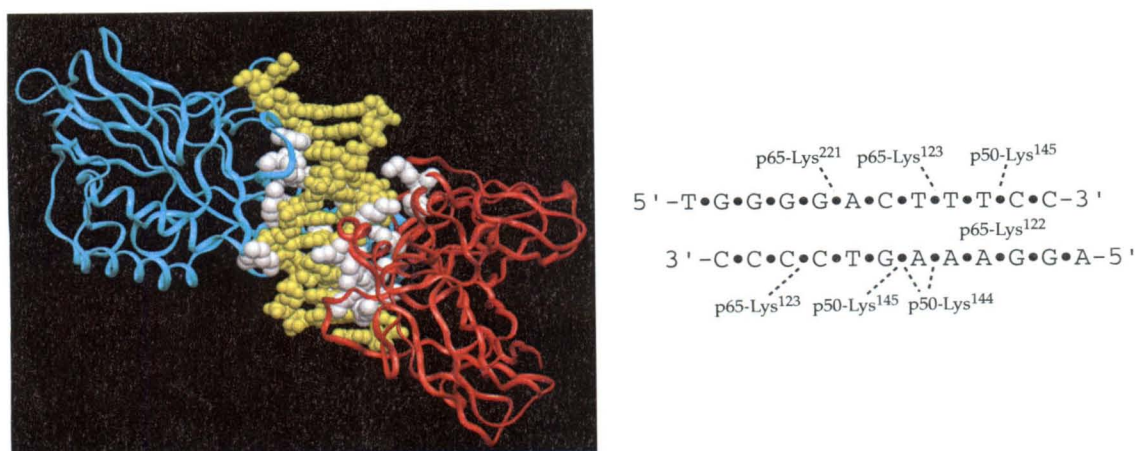


Figure 5.1. (left) Crystal Structure of p50p65 NF- κ B heterodimer:DNA complex. The p50 and p65 monomers are colored blue and red respectively. Amino acid side chains that contact DNA are colored white. (right) NF- κ B contacts to the minor groove and phosphate backbone contacts.

contacts and the minor groove is accessible with the exception of p50-Lys145, which contacts phosphates on either side of the minor groove of DNA (21) (see Figure 5.1).

We report here the synthesis and DNA binding properties of a series of hairpin polyamides targeted to the NF- κ B binding site and their ability to inhibit NF- κ B. Polyamides were designed to bind overlapping sites along the protein-binding site to determine which sites provided optimal inhibition. The DNA binding affinity and specificity of each of these compounds for their target DNA site was evaluated using quantitative DNase I footprinting. Gel mobility shift assays were employed to investigate the ability of these polyamides to inhibit the interaction of NF- κ B with DNA.

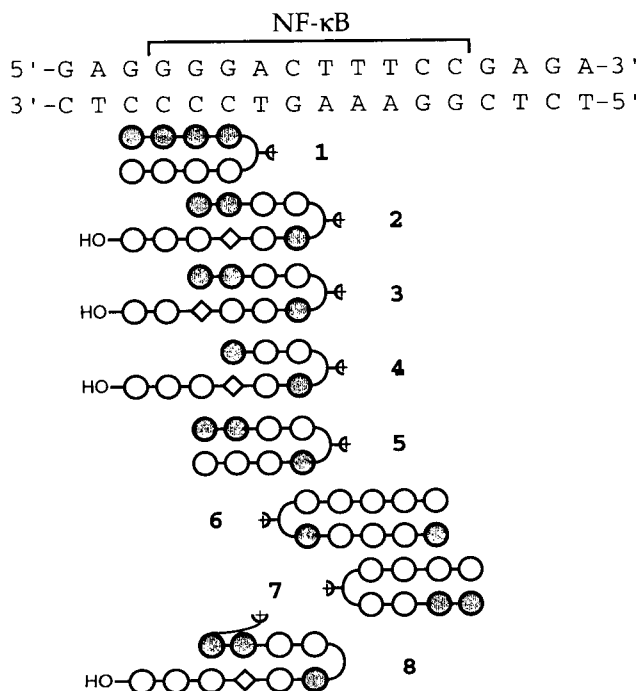


Figure 5.2. A ball and stick representation of polyamides bound to an NF- κ B DNA binding site. Filled circles represent imidazole rings while open circle denote pyrrole rings. The diamond represents β -alanine, OH represents a propanol amide and R-diaminobutyric acid is depicted as a curved line and a plus sign.

Results and Discussion

Polyamide Design and Synthesis. Polyamides **1-8** that cover the NF- κ B binding site were synthesized using solid-phase methods as previously described (22-24). Polyamides **1** and **5-7** were synthesized on oxime resin and cleaved with methyl amine to yield methyl amide tails, which allow the binding of sequences with G,C base pairs flanking the paired rings (24). Polyamides **2-4** and **8** were synthesized on Boc- β -alanine-PAM-resin followed by reductive cleavage to yield a β -propanol tail (22). Ring- β pairings were incorporated into these polyamides, since linkage 5 of contiguous rings results in decreased binding (25). Unpaired pyrrole rings were placed over two G,C base pairs because previous attempts to bind four consecutive G's with rings pairs results in low affinity compounds (26). Presumably the pyrrole ring should be positioned to avoid the exocyclic amine of guanine, since it is pushed up against the other side of the groove. Compound **8** incorporates a 1-(N-Boc-3-aminopropyl)-2-imidazole carboxylate monomer, allowing the charge to be located on the terminal imidazole. The monomer was synthesized in two steps from commercially available N-(3-aminopropyl)imidazole (Figure 5.3).

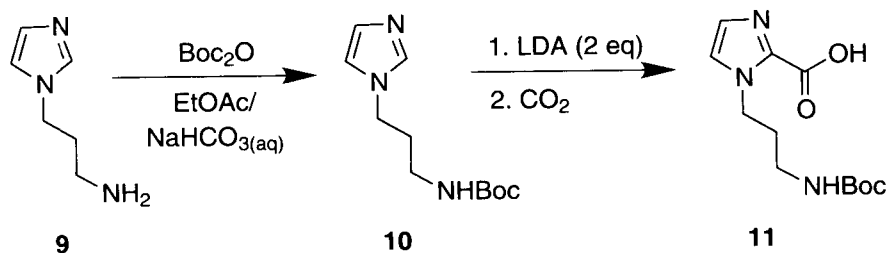


Figure 5.3. Synthesis of 1-(N-Boc-3-aminopropyl)-2-imidazole carboxylate.

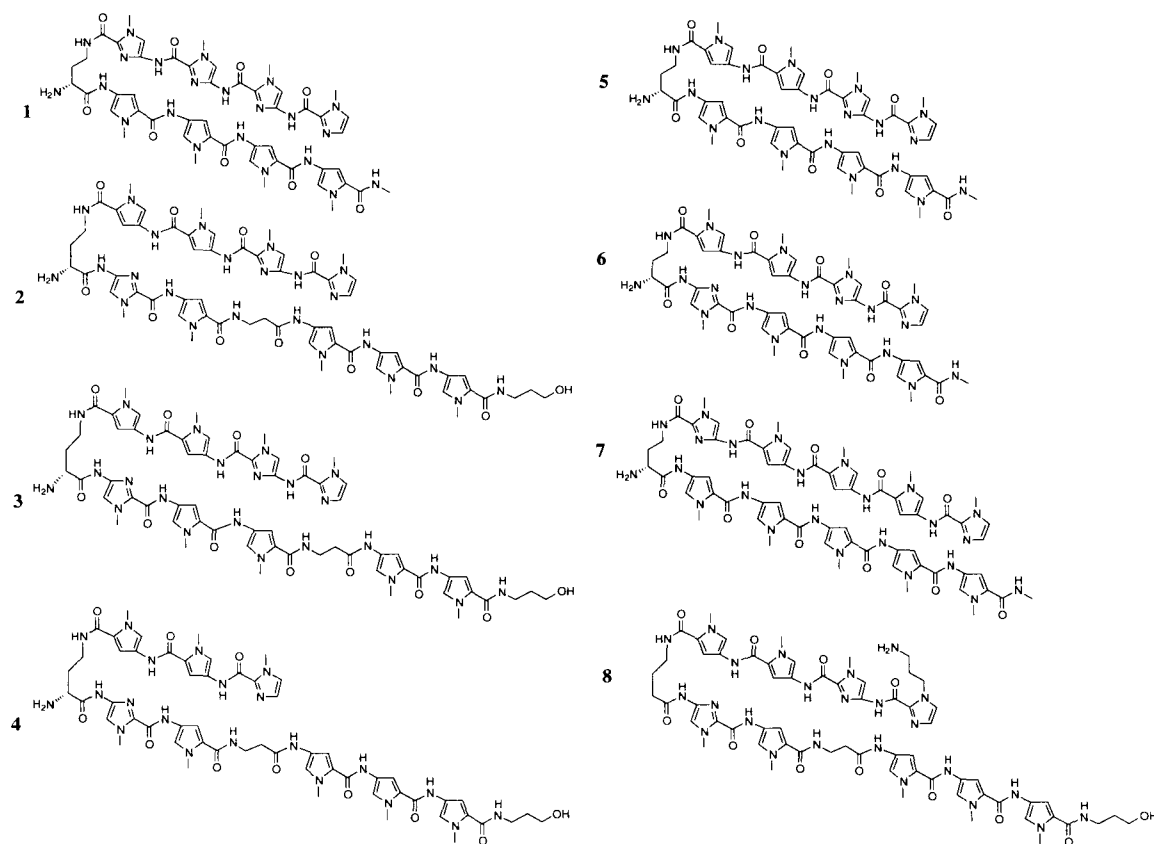


Figure 5.4. Structures of polyamides **1-8** targeted to the NF- κ B binding site.

Quantitative DNase I footprint titrations. Polyamide **1** binds its target site with relatively high affinity, but as with the parent compounds, coating is observed soon after occupation of the target site (30 nM). As a result the compound's lack of specificity will limit its utility. Series **2-4** reveals decreasing affinity for the target site. The pairing of β with the N-terminal imidazole monomer appears to be less optimal than an internal ring- β pair since polyamide **3** has a binding affinity fivefold lower than that of compound **2**. Compound **4** that has only two ring pairings and therefore shows lower affinity than either **2** or **3** and displays little specificity for the target sites, since many match sites exist on the labeled fragment. Evidently the unpaired pyrrole ring in each of these compounds is well tolerated by the G,C base pair. Presumably A,T base pairs would also be tolerated

by the ring. The unpaired pyrrole ring does allow for binding of the four consecutive G's with high affinity, which has not been accomplished previously. This motif will allow exploitation of the binding site immediately adjacent to the four G's for specificity, while still binding the G-rich site. Larger compounds were synthesized based on this motif in order to increase the specificity of the compounds by binding a larger binding site. Two imidazoles were incrementally added to the N-terminus of compound **2** to provide ring pairing across the G,C base pairs. Both compounds did not show binding of their target site and began to nonspecifically bind DNA at 30 nM. The eight-ring hairpin analogue of **2**, compound **5** binds with comparable affinity and specificity. Relocation of the charge from the turn to the terminal imidazole in compound **8** also did not affect binding affinity or specificity. The failure of the larger polyamides to bind DNA suggests that the structure of the DNA is irregular. Compounds **6** and **7** show good affinity and specificity for their target sites comparable to other hairpin polyamides that have been reported.

Table 5.2. Equilibrium Association Constants (M^{-1})^a

Polyamide	K_a
ImImImIm- ^{H₂N} γ-PyPyPyPy-NHCOMe (1)	8.0×10^9
ImImPyPy- ^{H₂N} γ-ImPy-β-PyPyPy-NHCOPrOH (2)	$9.3 (\pm 1.6) \times 10^9$
ImImPyPy- ^{H₂N} γ-ImPyPy-β-PyPy-NHCOPrOH (3)	$1.4 (\pm 0.1) \times 10^9$
ImPyPy- ^{H₂N} γ-ImPy-β-PyPyPy-NHCOPrOH (4)	$2.2 (\pm 0.6) \times 10^8$
ImImPyPy- ^{H₂N} γ-ImPyPyPy-NHCOMe (5)	$2.2 (\pm 0.3) \times 10^{10}$
ImPyPyPyIm- ^{H₂N} γ-PyPyPyPyPy-NHCOMe (6)	$6.4 (\pm 1.6) \times 10^9$
ImImPyPy- ^{H₂N} γ-PyPyPyPy-NHCOMe (7)	$1.3 (\pm 0.2) \times 10^9$
Im(C ₃ NH ₂)ImPyPy- ^{H₂N} γ-ImPy-β-PyPyPy-NHCOPrOH (8)	

^aValues reported are the mean values obtained from three DNase I footprint titration experiments except compounds, **1** and **8**, which are averages of two gels. The assays were carried out at 22 °C, 10 mM Tris•HCl (pH 7.0), 10 mM KCl, 10 mM MgCl₂, and 5 mM CaCl₂. The match sites are 5'-agAGGGGac-3' for **1**, 5'-gaGGGGACTtt-3' for **2-4** and **8**, ggGGACTtt-3' for **5**, 5'-ggACTTTCcg-3' for **6** and 5'-acTTTCCga-3' for **7**. Match sites are in upper case letters.

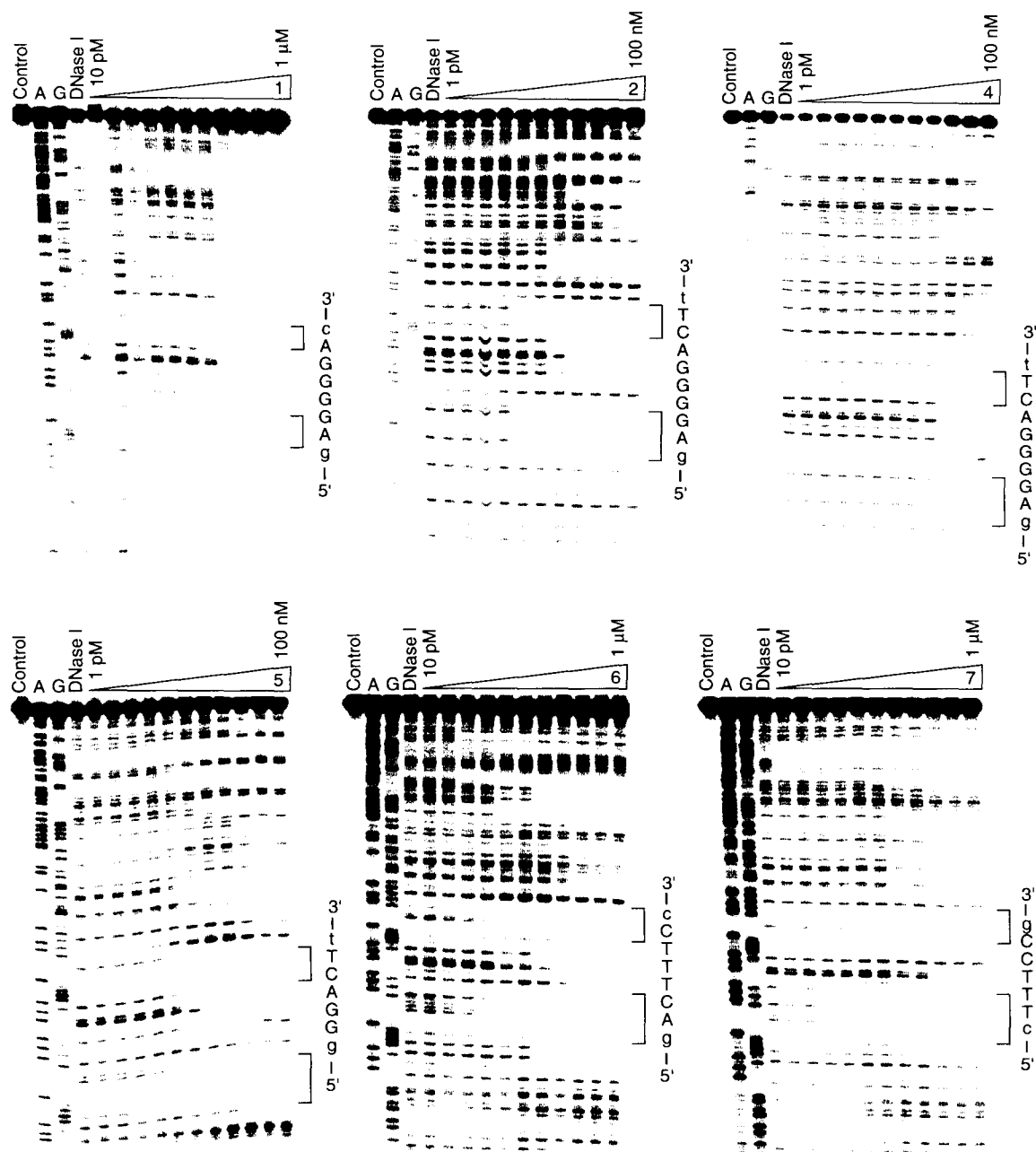


Figure 5.5. Storage phosphor autoradiograms of quantitative DNase I footprint titration experiments with ImImImIm-^{H2N} γ -PyPyPyPy-NHCOMe (1), ImImPyPy-^{H2N} γ -ImPy- β -PyPyPy-NHCOPROH (2), ImImPyPy-^{H2N} γ -ImPy- β -PyPyPy-NHCOPROH (3), ImPyPy-^{H2N} γ -ImPy- β -PyPyPy-NHCOPROH (4), ImImPyPy-^{H2N} γ -ImPyPyPy-NHCOMe (5), ImPyPyPyIm-^{H2N} γ -PyPyPyPyPy-NHCOMe (6), ImImPyPy-^{H2N} γ -PyPyPyPyPy-NHCOMe (7), Im(C₃NH₂)ImPyPy-^{H2N} γ -ImPy- β -PyPyPy-NHCOPROH (8) on the 3'-³²P-end-labeled restriction fragment. All reactions contained 10 kcpm restriction fragment, 10 mM Tris•HCl (pH 7.0), 10 mM KCl, 10 mM MgCl₂ and 5 mM CaCl₂ and were performed at 22 °C. For compounds 1, 6 or 7 Lane 1, intact DNA; lane 2, A-specific reaction; lane 3, G-specific reaction; lane 4, DNase I standard; lanes 5-15, 10 pM, 30 pM, 100 pM, 300 pM, 1 nM, 30 nM, 100 nM, 300 nM, 1 μ M. For compounds 2, 4 or 5 Lane 1, intact DNA; lane 2, A-specific reaction; lane 3, G-specific reaction; lane 4, DNase I standard; lanes 5-15, 1 pM, 30 pM, 10 pM, 30 pM, 100 pM, 300 pM, 1 nM, 30 nM, 100 nM. Each gel contains two match binding sites that are bracketed. Match binding sites are to the right of each gel in capital letters.

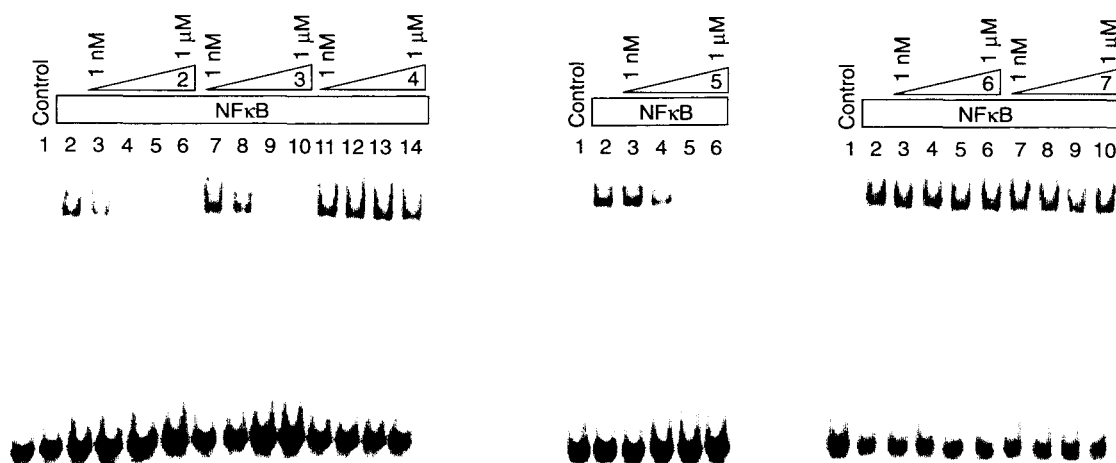


Figure 5.6. NF- κ B gel mobility shift assay with polyamides 1-7.

NF- κ B Binding Inhibition by Polyamides determined by Gel Shift.

Gel mobility shift assays were performed with P-32 labeled double-stranded oligonucleotides and nuclear extract isolated from TNF α activated 293T cells (17). Compounds **2**, **3** and **5** that bind the GGGAC subsite of the NF- κ B binding site inhibited the protein's binding to DNA at nanomolar concentrations. Higher concentrations were necessary for inhibition by **3**, a low affinity binder than compounds **2** or **5**, indicating that inhibition is dependent on the binding constant. Polyamides **6** and **7** bound the TTCC subsite of the NF- κ B binding site did not inhibit at concentrations up to 1 μ M. Inhibition at low nanomolar concentrations compare favorably to other protein binding inhibition results with polyamides.

Analysis of the NF- κ B DNA crystal structure reveals two possible mechanisms for inhibition of the complex formation by polyamides: allosteric and steric. The minor groove of the DNA bound by NF- κ B has a narrow minor groove of approximately 4 angstroms (21). Polyamides are known to slightly enlarge the minor groove when bound there to approximately 7.5 angstroms (27) (Figure 5.7A). The p50 half of the heterodimer

makes extensive major groove contacts on the 5'-GGGAC-3' subsite of the binding site, but few minor groove interactions in this region. Potentially the polyamide binds to the DNA sequence, enlarging the minor groove, stabilizing a conformation that cannot be bound by this half of the NF- κ B dimer. Alternatively, the p65-Lys221 in the minor groove in this region could be blocked by these polyamides (Figure 5.7B). Likely a combination of these interactions contributes to the inhibition of the p50p65 heterodimer.

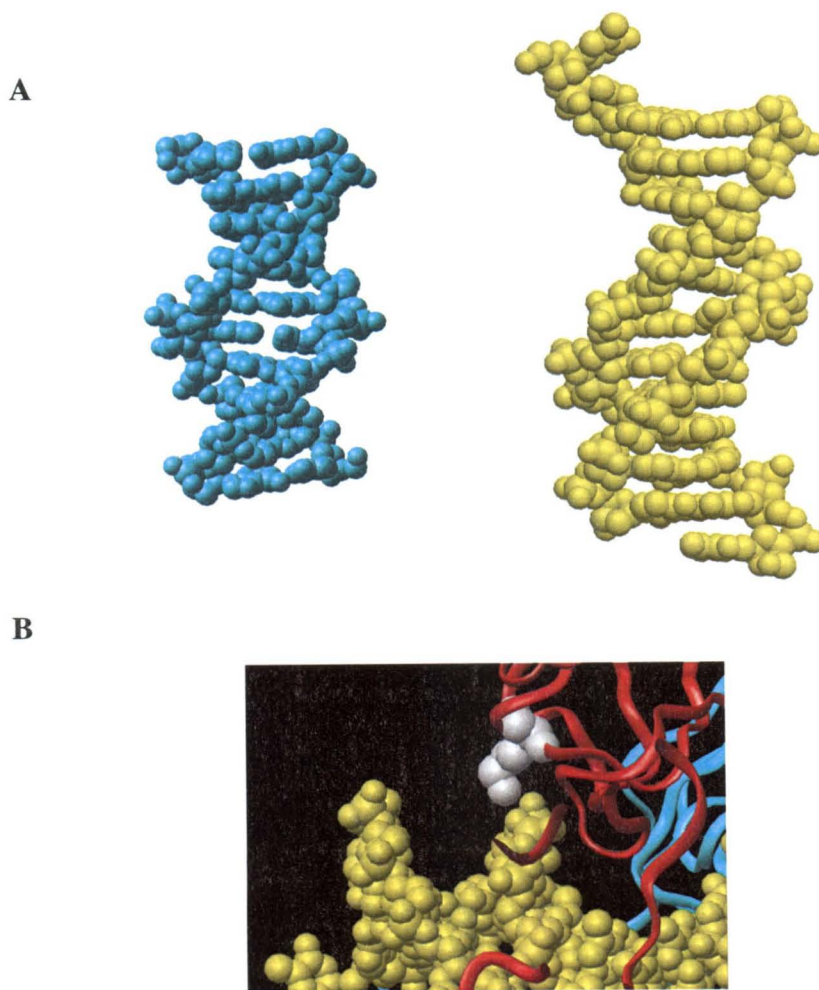


Figure 5.7. (A) Left, a structure of DNA when bound by a polyamide with a wide minor groove. Right, the structure of DNA when bound by NF- κ B with a narrow minor groove. (B) DNA minor groove contact by p65-Lys221.

Cell inhibition

Once it was established that polyamides could inhibit DNA binding by NF- κ B, we set out to determine whether the compounds could inhibit expression of a gene that used NF- κ B as a transcription factor. A plasmid with an NF- κ B responsive promoter attached to a luciferase reporter was transfected into 293T cells (28). Polyamides were added to the cell media to give a final concentration of 100 nM and 1 μ M and incubated overnight. TNF α was added to the cells to activate transcription. Little, if any, decreased transcription was observed for compounds **1-8** tested (any decrease observed was within the margin of error for the assay).

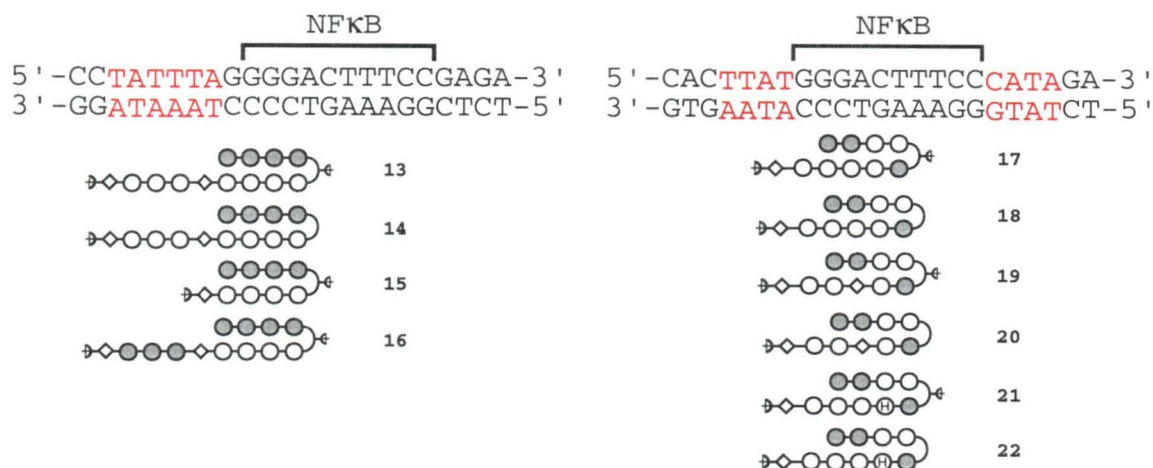


Figure 5.8. Additional polyamides synthesized to inhibit NF- κ B. Polyamides are represented as in Figure 2, except an H depicts a hydroxypyrrole ring. Red colored base pairs show site where the binding site was altered.

In parallel with this study, researchers in the Dervan group were studying cell permeability by polyamide-dye conjugates. Preliminary results indicated that the 2- β -2 polyamide motif was less cell permeable than a fully paired eight-ring hairpin polyamide. In light of this data, additional compounds **13-22** were synthesized that target the region that was known to inhibit NF- κ B-DNA complex formation (Figure 5.8). A variety of motifs were used in an effort to see if these compounds would have increased cellular

uptake. The reporter plasmid was mutated to allow binding of the different motifs. These compounds incorporated fully ring paired analogues, β -ring pairings, β -Dp tails, extended hairpins and Hp/Py pairings. Although these compounds except **16** inhibited NF- κ B DNA-binding, they were unsuccessful at downregulating the NF- κ B driven reporter in cell culture.

Conclusions

A series of polyamides were synthesized that target overlapping regions in a NF- κ B binding site. These compounds include a new motif that incorporates unpaired pyrrole rings over a C,G base pair that allow binding of four consecutive guanines with subnanomolar affinity as well as the first application of polyamides with methyl amide tails. This successful targeting of the 5'-CCTT-3' subsite with the polyamides with methyl amine tails demonstrates the utility of the new tail for recognition of G,C rich sequence that would be difficult using previous methods. Polyamides targeted to the left side of the protein binding site inhibited the NF- κ B-DNA complex at low nanomolar concentrations. The successful inhibition of NF- κ B suggests that polyamides could likely be used to inhibit other proteins that are members of the Rel homology family of DNA-binding proteins. Once cell trafficking of polyamides is figured out, then we hope to use these compounds to investigate their effects on gene regulation. One can envision inhibiting a subset of NF- κ B driven genes and observing the effect on biological systems.

Experimental

Monomer Synthesis.

1-(N-Boc-3-aminopropyl)imidazole. Boc anhydride (17.5 g, 80 mmol) was added to a solution of 1-(3-aminopropyl) imidazole in ethyl acetate (120 ml) and 20% aqueous potassium carbonate (100 ml). After stirring for 8 h at room temperature, the organic layer was separate, washed with sat'd sodium bicarbonate (1 × 50 ml), brine (1 × 50 ml), and concentration *in vacuo* to yield a colorless oil (15.1 g, 83.3% yield). ¹H NMR (DMSO-*d*₆): δ 7.61 (s, 1H), 7.12 (s, 1H), 6.92 (bt, 1H), 6.83 (s, 1H), 3.91 (t, 2H, J= Hz), 2.86 (d of d, 2H, J= Hz), 1.78 (m, 2H), 1.35 (s, 9H).

1-(N-Boc-3-aminopropyl)-2-imidazole carboxylate. A solution of 1-(N-Boc-3-aminopropyl)imidazole (13.8 g, 61.3 mmol) in dry THF (150 ml) under argon was cooled to -72 °C. nButLi (61 ml, 2.0 M in hexanes) was added to the reaction mixture. After stirring for 1.5 h, CO₂ (g) was bubbled through the reaction mixture for 2 h. The reaction was acidified with HCl (aq, 1 N, ~61 ml) to pH 8. The aqueous layer was frozen and lyophilized to yield a white solid (11.4 g). The crude solid was purified on reverse phase C18 silica gel with 0.1% TFA (aq) and acetonitrile to yield a white solid (5.7 g, 34.6% yield). ¹H NMR (DMSO-*d*₆): δ 7.16 (s, 1H), 7.01 (bt, 1H, J= Hz), 6.82 (s, 1H), 4.45 (t, 2H, J= Hz), 2.82 (d of d, 2H, J= Hz), 1.75 m (2H), 1.37 (s, 9H).

Polyamide Synthesis. As previously reported (2, 22, 24).

Quantitative DNase I Footprinting. As previously reported (29).

Gel Mobility Assays. As previously reported (17).

Acknowledgments

We are grateful to the National Institutes of Health for research support, Bristol-Myers Squibb and the Ralph M. Parson Foundation for predoctoral fellowships to N.R.W.

References

1. Dervan, P. B. (2001) *Bioorganic & Medicinal Chemistry* 9, 2215-2235.
2. Dervan, P. B., and Bürli, R. W. (1999) *Current Opinion in Chemical Biology* 3, 688-693.
3. White, S., Baird, E. E., and Dervan, P. B. (1997) *Chemistry & Biology* 4, 569-578.
4. White, S., Szewczyk, J. W., Turner, J. M., Baird, E. E., and Dervan, P. B. (1998) *Nature* 391, 468-471.
5. Gottesfeld, J. M., Turner, J. M., and Dervan, P. B. (2000) *Gene Expression* 9, 77-91.
6. Gottesfeld, J. M., Neely, L., Trauger, J. W., Baird, E. E., and Dervan, P. B. (1997) *Nature* 387, 202-205.
7. Neely, L., Trauger, J. W., Baird, E. E., Dervan, P. B., and Gottesfeld, J. M. (1997) *Journal of Molecular Biology* 274, 439-445.
8. Dickinson, L. A., Gulizia, R. J., Trauger, J. W., Baird, E. E., Mosier, D. E., Gottesfeld, J. M., and Dervan, P. B. (1998) *Proceedings of the National Academy of Sciences of the United States of America* 95, 12890-12895.
9. Chiang, S. Y., Burli, R. W., Benz, C. C., Gawron, L., Scott, G. K., Dervan, P. B., and Beerman, T. A. (2000) *Journal of Biological Chemistry* 275, 24246-24254.
10. Bremer, R. E., Baird, E. E., and Dervan, P. B. (1998) *Chemistry & Biology* 5, 119-133.
11. Bremer, R. E., Wurtz, N. R., Szewczyk, J. W., and Dervan, P. B. (2001) *Bioorganic & Medicinal Chemistry* 9, 2093-2103.

12. Sen, R., and Baltimore, D. (1986) *Cell* 47, 921-928.
13. Baldwin, A. S. (1996) *Annual Review of Immunology* 14, 649-683.
14. Verma, I. M., Stevenson, J. K., Schwarz, E. M., Vanantwerp, D., and Miyamoto, S. (1995) *Genes & Development* 9, 2723-2735.
15. Beg, A. A., and Baltimore, D. (1996) *Science* 274, 782-784.
16. Liu, Z. G., Hsu, H. L., Goeddel, D. V., and Karin, M. (1996) *Cell* 87, 565-576.
17. VanAntwerp, D. J., Martin, S. J., Kafri, T., Green, D. R., and Verma, I. M. (1996) *Science* 274, 787-789.
18. Wang, C. Y., Mayo, M. W., and Baldwin, A. S. (1996) *Science* 274, 784-787.
19. Siebenlist, U., Franzoso, G., and Brown, K. (1994) *Annual Review of Cell Biology* 10, 405-455.
20. Ghosh, S., Gifford, A. M., Riviere, L. R., Tempst, P., Nolan, G. P., and Baltimore, D. (1990) *Cell* 62, 1019-1029.
21. Chen, F. E., Huang, D. B., Chen, Y. Q., and Ghosh, G. (1998) *Nature* 391, 410-413.
22. Herman, D. M., Baird, E. E., and Dervan, P. B. (1998) *Journal of the American Chemical Society* 120, 1382-1391.
23. Baird, E. E., and Dervan, P. B. (1996) *Journal of the American Chemical Society* 118, 6141-6146.
24. Belitsky, J. B., Nguyen, D. H., and Wurtz, N. R. (2001) *Bioorganic & Medicinal Chemistry*, submitted.
25. Kelly, J. J., Baird, E. E., and Dervan, P. B. (1996) *Proceedings of the National Academy of Sciences of the United States of America* 93, 6981-6985.

26. Swalley, S. E., Baird, E. E., and Dervan, P. B. (1997) *Journal of the American Chemical Society* 119, 6953-6961.
27. Kielkopf, C. L., Baird, E. E., Dervan, P. D., and Rees, D. C. (1998) *Nature Structural Biology* 5, 104-109.
28. Pomerantz, J. L., and Baltimore, D. (1999) *Embo Journal* 18, 6694-6704.
29. Trauger, J. W., and Dervan, P.B. (2001) *Methods in Enzymology* 340, 450-466.

CHAPTER 6

Sequence Specific Alkylation of DNA by Hairpin Pyrrole-Imidazole Polyamide Conjugates

The text of this chapter was taken in part from a publication coauthored with Prof. Peter Dervan.

(N.R. Wurtz & P.B. Dervan *Chem. & Biol.*, **2000**, 7, 152-161.)

Abstract

Pyrrole/imidazole polyamides are synthetic ligands that recognize predetermined sequences in the minor groove of DNA with affinities and specificities comparable to DNA-binding proteins. As a result of their DNA binding properties, polyamides could deliver reactive moieties for covalent reaction at specific DNA sequences and thereby inhibit DNA-protein interactions. Site-specific alkylation of DNA could be a useful tool for regulation of gene expression. As a minimal first step, we set out to design and synthesize a class of hairpin polyamides equipped with DNA alkylating agents and characterize the specificity and yield of covalent modification. Bis(dichloroethylamino)-benzene derivatives of the well-characterized chlorambucil (CHL) were attached to the γ -turn of an eight-ring hairpin polyamide targeted to the HIV promoter. We found that a hairpin polyamide-CHL conjugate binds and selectively alkylates predetermined sites in the HIV promoter at subnanomolar concentrations. Cleavage sites were determined on both strands of a restriction fragment containing the HIV-1 promoter revealing good specificity and high yield of alkylation. The ability of polyamide-CHL conjugates to sequence specifically alkylate double stranded DNA in high yield and at low concentration sets the stage for testing their use as regulators of gene expression in cell culture and ultimately in complex organisms.

Introduction

Small molecules specifically targeted to any predetermined DNA sequence would be useful tools in molecular biology and potentially in human medicine. Polyamides containing pyrrole (Py), and imidazole (Im) amino acids, are synthetic ligands that have an affinity and specificity for DNA comparable to naturally occurring DNA binding proteins (1-5). DNA recognition depends on side-by-side amino acid pairings in the minor groove. An antiparallel pairing of Im opposite Py (Im/Py) distinguishes G•C from C•G and both of these from A•T/T•A base pairs (6-9). A Hp/Py pair specifies T•A from A•T and both of these from C•G/G•C (10, 11). The linker amino acid, γ -aminobutyric acid (γ), connects polyamide subunits C \rightarrow N in a “hairpin motif,” and these ligands bind to predetermined target sites with >100 fold enhanced affinity relative to unlinked dimers (1-4, 12). Paired β -alanine residues (β/β) restore the curvature of the dimer for recognition of larger binding sites and, in addition, code for AT/TA base pairs (13-15). Eight-ring hairpin polyamides permeate eukaryotic cells and have been shown to regulate transcription by targeting promoters of specific genes (16, 17).

Sequence specific covalent attachment in the DNA minor groove by small molecules

In unpublished observations, hairpin polyamides do not appear to inhibit gene expression when bound to the coding region of genes. In order to inhibit elongation by the RNA polymerase enzymes, it may be that one needs either higher affinity class of DNA binders or alternatively bifunctional molecules that covalently attach to predetermined sequences in the minor groove of DNA. In this paper we address this latter class of molecules, the design of hairpin polyamides with separate domains for

DNA binding and DNA covalent attachment. This is a minimal first step toward creating new research tools for the field of functional genomics.

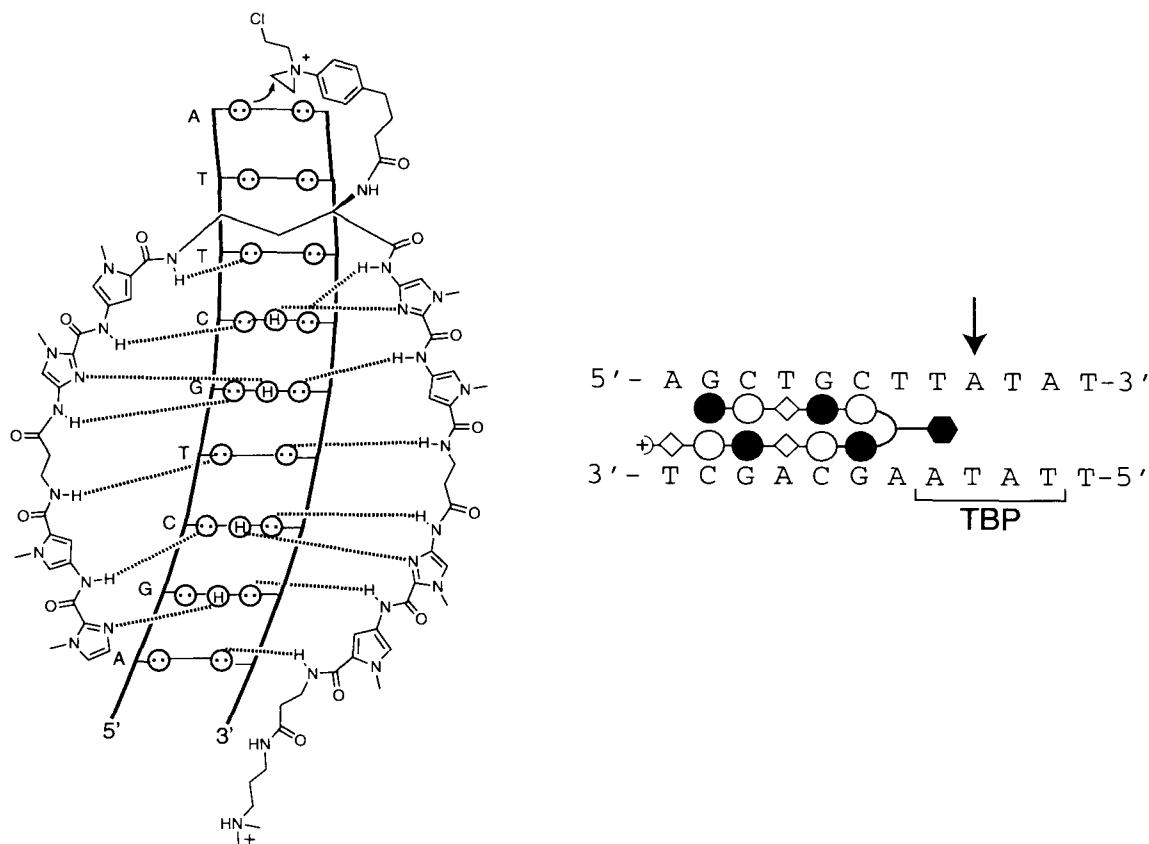


Figure 6.1. Top, the hydrogen-bonding model and alkylation mechanism of polyamide-chlorambucil conjugate ImPy- β -ImPy-(R)^{CHL}- γ -ImPy- β -ImPy- β -Dp (**2**) bound to the minor groove of 5'-AGCTGCT-3'. Circles with two dots represent the lone pairs of N3 purines and O2 of pyrimidines. Circles containing an H represent the N2 hydrogens of guanines. Putative hydrogen bonds are illustrated by dotted lines. Bottom, model of polyamide conjugate **2** bound to the match site 5'-AGCTGCT-3'. Black and white circles represent imidazole (Im) and pyrrole (Py) polyamide rings, respectively. Diamonds and hexagons represent β -alanine (β) and chlorambucil (CHL) respectively. (R)-2,4-diaminobutyric acid ((R) γ) and dimethylamino-propylamide (Dp) are depicted as a curved line and a plus sign, respectively.

Hairpin polyamide-bis(dichloroethylamino)benzene conjugates

Nitrogen mustards are well characterized DNA alkylators with little sequence specificity (18). A freely diffusing bis(dichloroethylamino)benzene moiety such as chlorambucil reacts with the more nucleophilic N7 of guanine in the major groove, but will alkylate the N3 of adenine in the minor groove when attached to minor groove DNA

binding ligands (19, 20). The combination of subnanomolar binding hairpin polyamides and bis(dichloroethylamino)benzene derivatives creates a class of bifunctional agents that bind to predetermined DNA sequences with high affinity and specificity for subsequent covalent reactions at N3 of adenine. Key design issues are sites of attachment of the functional domains and the choice of linker length and flexibility. In order to target the adenines adjacent to polyamide binding sites, we chose to attach bis(dichloroethylamino)benzene derivatives to the chiral α -amino group on the γ -turn of a pyrrole/imidazole hairpin polyamide.

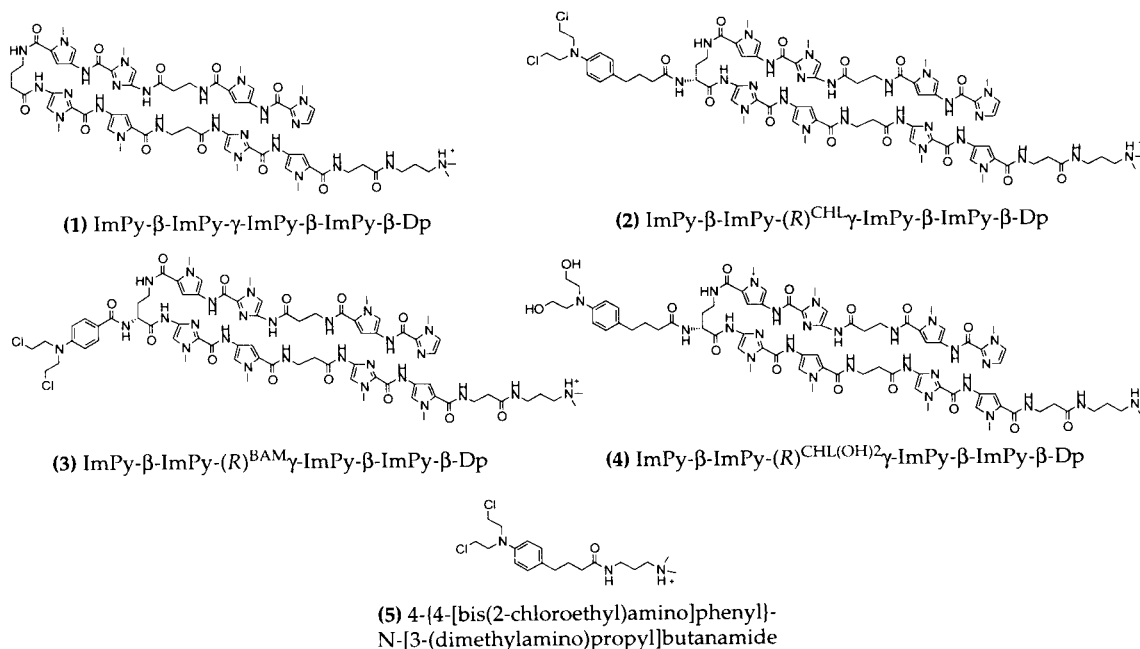


Figure 6.2. Structures of polyamides 1-4 and a nonspecific alkylator 5.

Polyamide 1 has been shown to bind the sequence 5'-WGCWGCW-3' adjacent to both sides of the TATA box in the HIV promoter with high affinity ($K_a = 2.0 \times 10^{10} \text{ M}^{-1}$) (16). Substitutions at the α position of the γ -turn using (R)-2,4-diaminobutyric acid potentially have only a modest effect on DNA binding affinity or specificity (21). We

chose this as the point of attachment of a nondiffusible electrophile that would react at the N3 of adenine adjacent to the polyamide-binding site. Chlorambucil (CHL), 4-[bis(2-chloroethyl)amino]benzene butanoic acid, which has a simple flexible linker and the less reactive 4-[bis(2-chloroethyl)amino]benzoic acid were chosen as alkylators. Since alkylation by the electrophile results in irreversible binding, a non-alkylating conjugate **4** in which the chlorides have been replaced with hydroxyls was also synthesized as a control to investigate energetic penalty (or lack thereof) on polyamide binding affinity and specificity for the target sequence. In addition compound **5** will serve as a control to compare alkylation of the alkylating moiety unlinked to a polyamide.

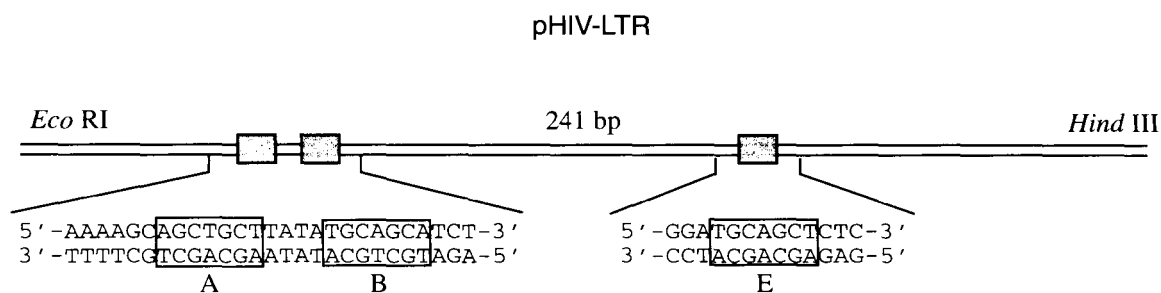


Figure 6.3. An illustration of the 241 bp pHIV-LTR *EcoRI/HindIII* restriction fragment with the position of the sequences indicated. The binding sites 5'-AGCTGCT-3', 5'-TGCAGCA-3' and 5'-TGCAGCT-3' are highlighted in gray.

Here, we report preliminary results on our first generation of hairpin polyamides that selectively alkylate and cleave DNA in the minor groove. Alkylation adducts and cleavage yields were determined using thermal cleavage assays on a 241 base pair restriction fragment and 120 base pair synthetic oligonucleotides. DNA binding affinity and specificity were determined by DNase I footprinting of the non-reactive conjugate **3**.

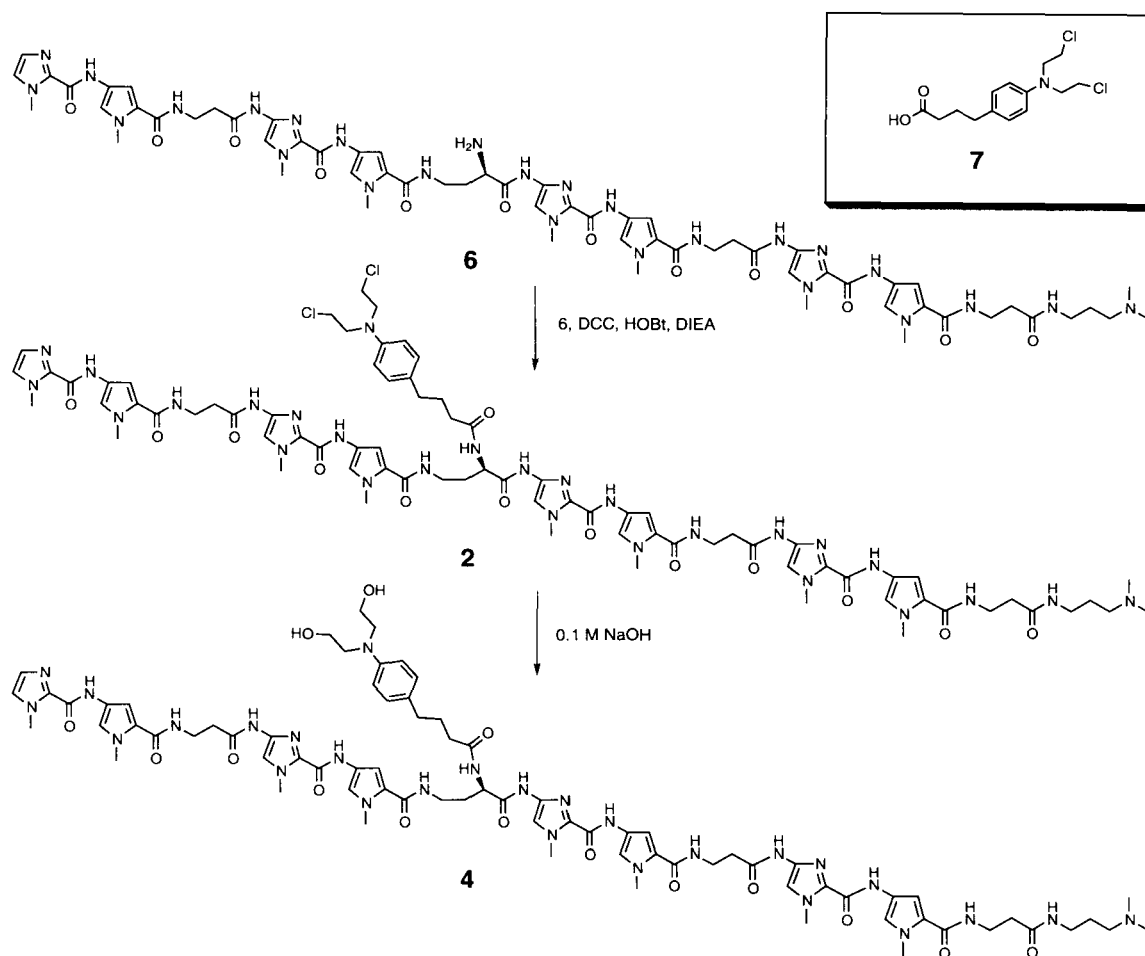


Figure 6.4. Synthetic scheme for preparation of polyamide conjugate, ImPy- β -ImPy-(R)^{CHL γ} -ImPy- β -ImPy- β -Dp (**2**). The HOBt-activated nitrogen mustard **7** is coupled with standard DCC/HOBt activation to polyamide **6** and purified by reverse phase HPLC (step i). The hydrolyzed conjugate **4** is prepared by addition of 0.1 M NaOH (step ii).

Results and Discussion

Synthesis of Conjugates

Polyamide **6**, ImPy- β -ImPy-(R)^{H₂N γ} -ImPy- β -ImPy- β -Dp, was prepared by manual solid phase polyamide synthesis (22). After purification by reverse phase high performance liquid chromatography (HPLC), the appropriate activated carboxylate derivative of bis(dichloroethylamino)benzene was coupled to the α -amino group on the γ -turn, using standard DCC/ HOBt conditions, to yield polyamide-nitrogen mustard

conjugates **2** and **3** (Figure 4). Control conjugate **4** was synthesized by allowing **2** to react with 0.1 M NaOH, followed by neutralization and lyophilization. Control compound **5** was synthesized by coupling chlorambucil to (dimethylamino)-propylamine (Dp). All compounds were purified by reverse phase HPLC. MALDI-TOF mass spectrometry analysis of each compound was consistent with the calculated mass of the compounds.

Polyamide binding affinity and specificity is unaltered by substitution

Quantitative DNase I footprint titration experiments were performed on compound **4** to measure the equilibrium association constant for the match binding sites (5'-WGCWGCW-3') on a 241 bp restriction fragment derived from the HIV-1 promoter region (Figure 6.5). The conjugate binds the match sites 5'-AGCTGCT -3' and 5'-TGCAGCA-3' with equilibrium association constants of $K_a = 1.6 \pm 0.7 \times 10^{10} \text{ M}^{-1}$ and $1.3 \pm 0.7 \times 10^{10} \text{ M}^{-1}$, respectively. This is remarkably similar to unsubstituted polyamide **1** ($K_a = 2.0 \times 10^{10} \text{ M}^{-1}$)(17). Affinities for the double base pair mismatch sites, 5'-TGTCGCC-3' and 5'-AGCAGCTC-3', were determined to be >100 fold lower than for the match sites. Attachment of the benzene moiety to the γ turn does not appear to affect the DNA binding affinity or specificity of the polyamide.

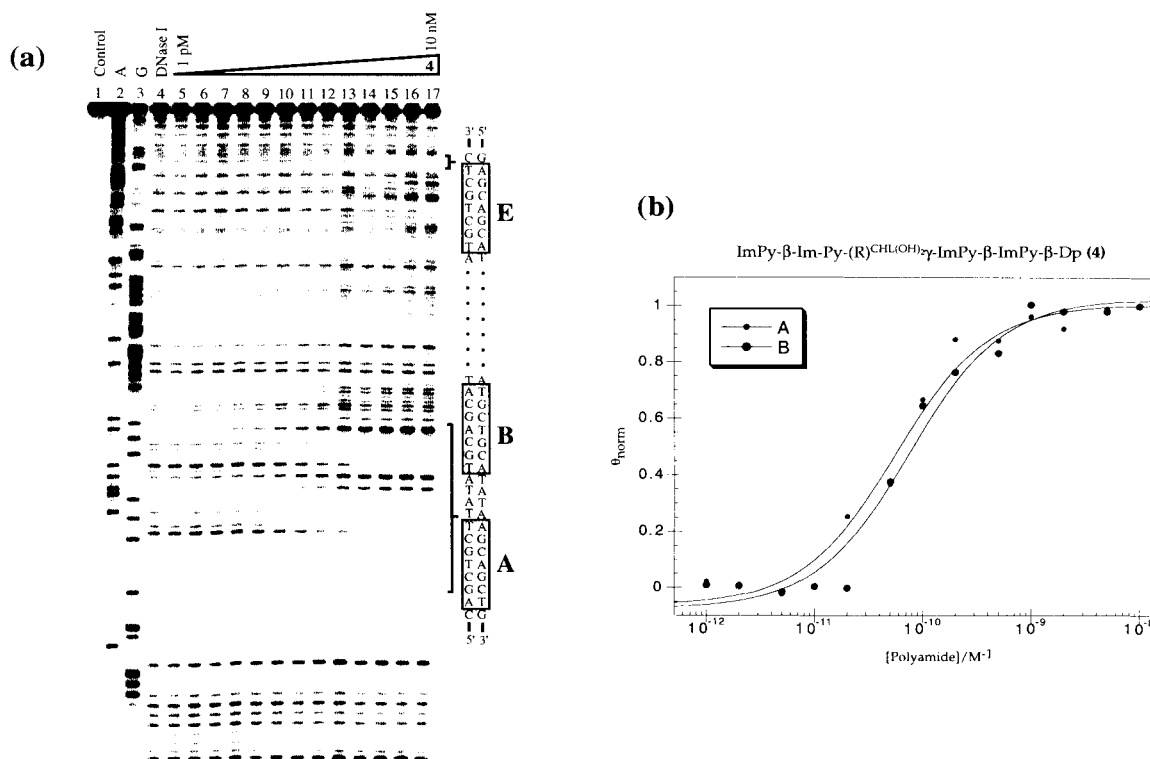


Figure 6.5 (a) Quantitative DNase I footprint titration experiments with ImPy- β -ImPy-(R)^{CHL(OH)₂}- γ -ImPy- β -ImPy- β -Dp (4) on the 3'-³²P-end-labeled 241 base pair *EcoRI/HindIII* restriction fragment from plasmid pHIV-LTR(23). Left, storage phosphor autoradiograms of 8% denaturing polyacrylamide gels used to separate the fragments generated by DNase I digestion. All reactions contained 20,000 cpm restriction fragment, 10 mM Tris•HCl (pH 7.0), 10 mM KCl, 10 mM MgCl₂ and 5 mM CaCl₂ and were performed at 22 °C. Lane 1, intact DNA; lane 2, A-specific reaction; lane 3, G-specific reaction; lane 4, DNase I standard; lanes 5-17, 1 pM, 2 pM, 5 pM, 10 pM, 20 pM, 50 pM, 100 pM, 200 pM, 500 pM, 1 nM, 2 nM, 5 nM, 10 nM, respectively. The 5'-AGCTGCT-3' (A), 5'-TGCAGCA-3' (B), and 5'-TGCTGCT-3' (E) match sites are shown in boxes on the right side of the autoradiogram. (b) Data from quantitative DNase I footprint titration experiments for ImPy- β -ImPy-(R)^{CHL(OH)₂}- γ -ImPy- β -ImPy- β -Dp (3) binding to the 5'-AGCTGCT-3' (A) and 5'-TGCAGCA-3' sites (B). θ_{norm} points were obtained using storage phosphor autoradiography and processed as described in the experimental section. The solid curves are best-fit Langmuir binding titration isotherms obtained from a nonlinear least-squares algorithm eq 2 ($n=1$).

DNA alkylation by polyamide conjugate proceeds in high yield

In order to measure DNA cleavage, thermal cleavage assays were performed on 3' and 5' ³²P labeled 241 base pair restriction fragments containing the HIV-1 promoter (Figure 6.6) (24). Alkylation in high yield was observed at subnanomolar concentrations for conjugate 2. In fact quantitative cleavage by conjugate 2 at 10 nM concentration was observed on one strand (bottom strand, Figure 6.2). In contrast, no cleavage was observed

at 10 μM with compound **3** or at 10 mM concentration of the freely diffusing chlorambucil-(dimethylamino) propylamine derivative **4**.

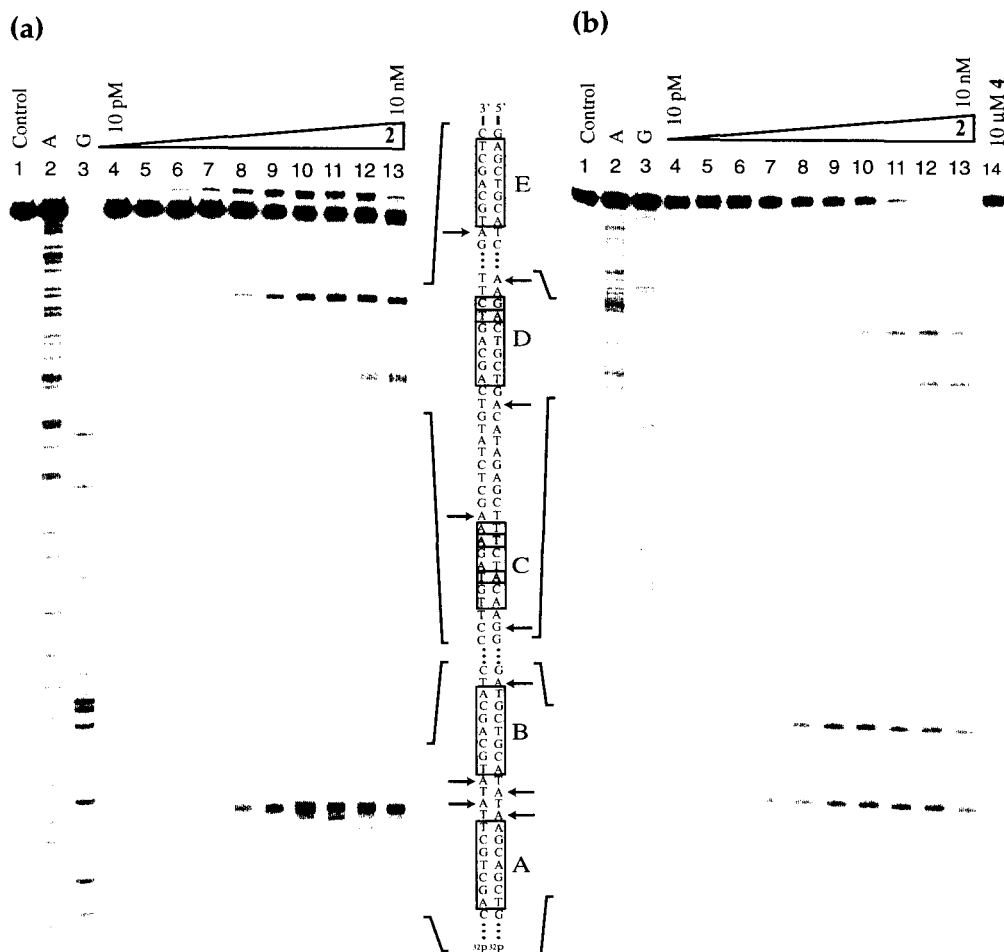


Figure 6.6. Thermal cleavage assay experiments with ImPy- β -ImPy-(R)^{CHL}-ImPy- β -ImPy- β -Dp (**2**) on the 5' and 3'-³²P-end-labeled 241 base pair *EcoRI/HindIII* restriction fragment from plasmid pHIV-LTR(23). Storage phosphor autoradiograms of 8% denaturing polyacrylamide gels used to separate the fragments generated by heat induced DNA cleavage at alkylation sites. All reactions contained 10,000 cpm restriction fragment, 10 mM Tris•HCl (pH 7.0), 10 mM KCl, 10 mM MgCl₂ and 5 mM CaCl₂ and were performed at 37° C. Following 24 hour of equilibration, the DNA pellet was resuspended in sodium citrate buffer (pH=7.2) and heated to 90° C for 15 min to thermally cleave at sites of adenine- or guanine-N3 lesions. (a) 5'-³²P-end-labeled restriction fragment. (b) 3'-³²P-end-labeled restriction fragment. (a-b) Lane 1, intact DNA; lane 2, A-specific reaction; lane 3, G-specific reaction; lanes 4-13, 10 pM, 20 pM, 50 pM, 100 pM, 200 pM, 500 pM, 1 nM, 2 nM, 5 nM, 10 nM, respectively. (b) Lane 14, 10 μM CHL-Dp (**5**). Center, alkylation sites on the DNA fragment, indicated by arrows. Sites 5'-TCGTCGACGA-3' (**A**), 5'-ACGTCGA-3' (**B**), and 5'-ACGTCGA-3' (**E**) are match sites and 5'-ACATCTT-3' (**C**) and 5'-TCGTCAG-3' (**D**) are double base pair mismatches (mismatches indicated by gray boxes).

Polyamide conjugate specifically alkylates adenines adjacent to binding sites

Upon heat-induced cleavage, all alkylation sites observed for conjugate **2** at subnanomolar concentrations were adjacent to target polyamide-binding sites (Figure 6.7). Due to the symmetry of the 5'-WGCWGCW-3' binding site, the polyamide binds in two orientations, alkylating adenine residues adjacent on either side of the polyamide binding site (Figure 7.7B). At concentrations greater than 1 nM, minor cleavage was observed proximal to double base pair mismatch sites as well. Specificity for reaction at match sites over mismatch sites is approximately 20-fold.

For the thermal cleavage assays, alkylation of adenine was observed at all but one interesting exception. The exception was cleavage of guanine at one of the mismatch sites. Reaction at N3 of guanines has previously been observed with tallimustine, a Py_3 conjugate with a bis(dichloroethylamino)benzene derivative (25, 26).

Effects of Tether on Alkylation

No alkylation was observed for benzoic acid mustard analogue **3** at concentrations up to 10 μM . This suggests that the short, rigid carbamido linker prevents alkylation for two possible reasons: Either the linker points the alkylator into the minor groove, severely decreasing the binding affinity of the conjugate or the alkylator is merely restricted from being close enough to the N-3 of adenine to allow for covalent attachment. This contrasts the longer, more flexible C-4 linker of chlorambucil, which enables the mustard moiety to alkylate base pairs adjacent to the polyamide match site.

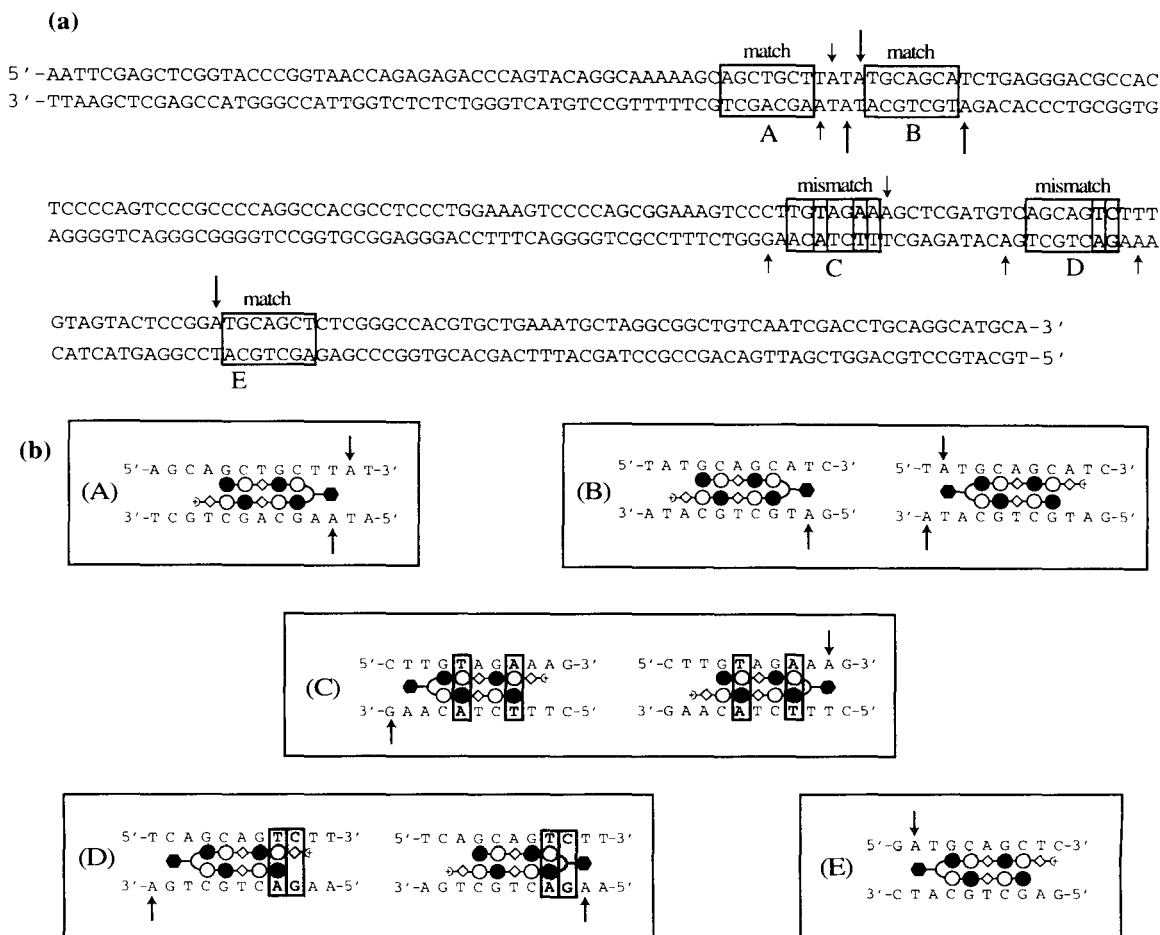


Figure 6.7. (a) Sequence of the 241 bp pHIV-LTR *Eco* RI/*Hind* III restriction fragment. The five binding sites are indicated by boxes. Base pair mismatches are indicated by gray boxes. Cleavage sites of adenine- or guanine-N3 lesions at 10 nM are indicated by arrows. Large and small arrows indicate major and minor alkylation sites, respectively. (b) Binding models for alkylation are shown. Models are colored as in Figure 1.

Alkylation of the exocyclic amine of guanine by polyamide conjugates

When the thermal cleavage assays were carried out on the 5'-end labeled restriction fragment, a slower mobility fragment appears in the gel electrophoresis assay above the intact DNA indicating the presence of a stable adduct retarding the full-length fragment. The stability of the adduct to heat and piperidine workups suggested that the adduct was the result of alkylation of the exocyclic amine of guanine in the minor groove. To test this, inosine substitution at that position was employed.

Inosine which lacks the exocyclic NH_2 of guanine was incorporated into a 120 base pair fragment at the putative guanine reaction site located adjacent to the match site. The strands with inosine substitutions, as well as a strand with no inosine substitutions, were 5' end labeled and annealed to a complementary strand and thermal cleavage assays were performed (Figure 6.8). The results of the thermal cleavage assay of the 5' labeled synthetic oligonucleotides without inosine substitutions were identical to alkylation of the restriction fragment (Figure 6.8A). Substitution of the guanine in the sequence 5'-AGCAGCTGCT-3' eliminated the anomalous band which results from the stable adduct (Figure 6.8B). Thus, certain DNA sequences allow the covalent reaction of the exocyclic amine of guanine by the polyamide-bis(dichloroethylamino)benzene conjugate.

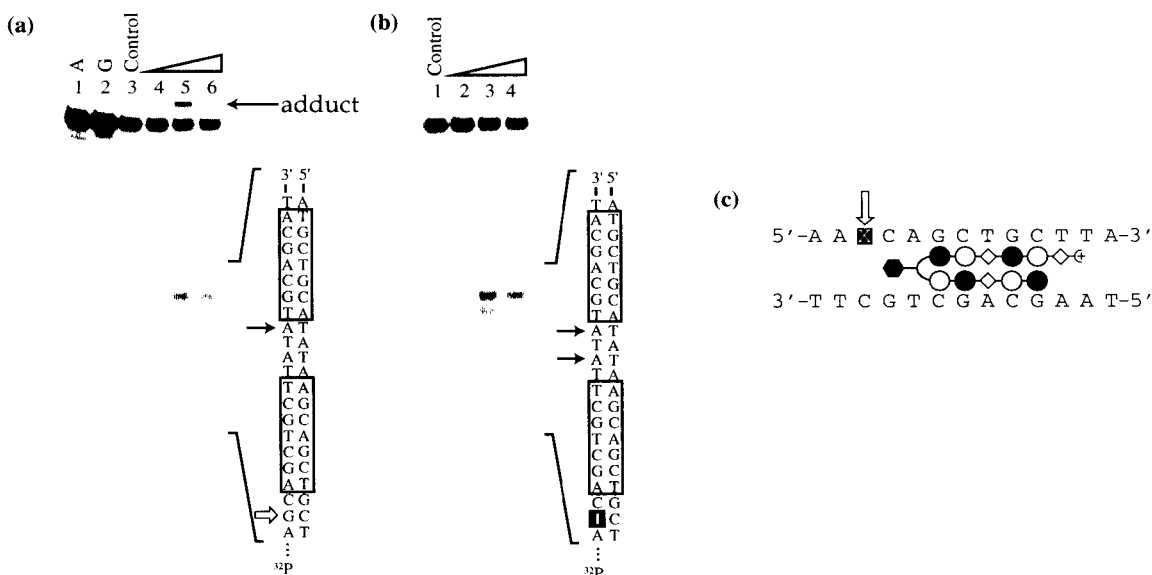


Figure 6.8. Thermal cleavage assay experiments with ImPy- β -ImPy-(R)^{CHL}- γ -ImPy- β -ImPy- β -Dp (**2**) on 5'-³²P-end-labeled 120 base pair oligonucleotides with and without inosine substitutions. **(a-b)** The solid arrows indicate cleavage bands from cleavage sites of adenine-N3 lesions on the restriction fragment. The hollow arrow indicates alkylation at the exocyclic amine of guanine. **(a)** The fragment contains the site 5'-AGCAGCTGCT. Lane 1, A-specific reaction; lane 2, G-specific reaction; lanes 3-6, 100 pM, 1 nM, 10 nM, respectively. **(b)** The fragments contain the site 5'-AICAGCTGCT-3'. Lanes 1-4, 100 pM, 1 nM, 10 nM, respectively. **(c)** Model bound to match site 5'-AGCTGCT-3'. X indicates site of inosine substitution. Models are colored as in Figure 1.

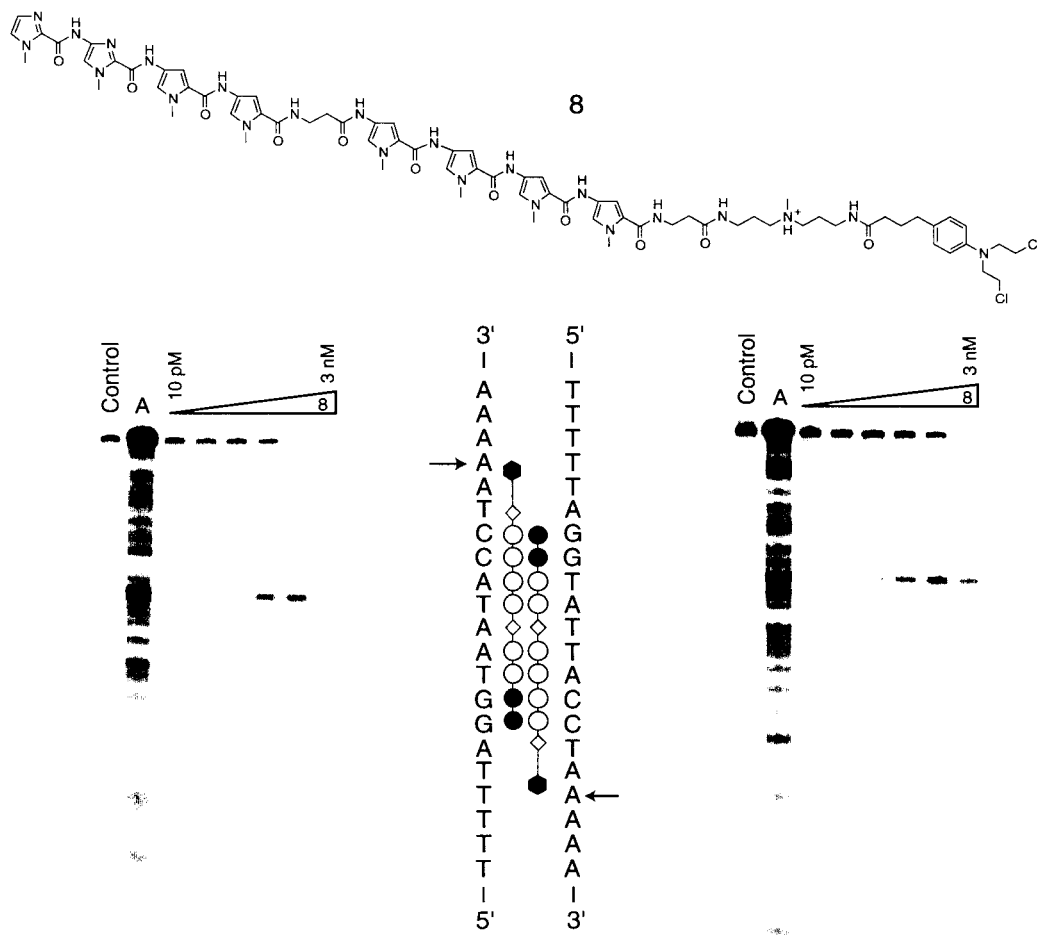


Figure 6.10. Alkylation by polyamide homodimer-chlorambucil conjugate.

Subsequent to this initial study of chlorambucil-polyamide conjugates, several other hairpin conjugates have been synthesized (data not shown). Good alkylation yields were observed 1-2 bp adjacent to match sites, confirming the generality of this design for sequence specific alkylation. Chlorambucil has also been attached to other polyamide motifs, including tandems and homodimers. Conjugate **8** displayed particularly good specificity and yield of alkylation when targeted to its match site on a restriction fragment. Quantitative cleavage was observed at 3 nM and cleavage was only observed at the match sites on either strand at these concentrations.

Conclusion

At subnanomolar concentrations the hairpin polyamide delivers the electrophilic bis(dichloroethylamino)benzene to the predetermined sequence and produces a specific covalent reaction in high yield in the minor groove of DNA. The attachment of the chlorambucil moiety at the γ -turn results in no significant alteration of the polyamide-DNA binding affinity. The parent eight-ring hairpin polyamide is cell permeable, and the next technical hurdle is to determine whether the new polyamide-bis(dichloroethylamino)benzene conjugates will also permeate eukaryotic cells. The parent hairpin polyamide **1** has been shown to inhibit HIV replication in cell culture, and this will form the basis of our next experiment. Conjugate **2** targeted to the same HIV-1 promoter can now be tested in cell culture, to determine its ability to inhibit transcription. Further downstream this should set the stage for site-specific alkylation in "coding region" of genes in order to inhibit transcription during the elongation phase. Sequence-specific DNA cleavage by multiple polyamide conjugates could provide a basis for "genetic microsurgery," whereby undesired gene segments or integrated viral DNAs could be selectively removed from a host genome *in vivo*.

Experimental

Materials

^1H NMR spectra were recorded on a General Electric-QE NMR spectrometer at 300 MHz with chemical shifts reported in parts per million relative to residual solvent.

UV spectra were measured in water on a Hewlett-Packard Model 8452A diode array spectrometer. Matrix assisted, laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF) was performed at the Protein and Peptide Microanalytical Facility at the California Institute of Technology. HPLC analysis was performed using a Beckman Gold Nouveau system using a Rainen C18, Microsorb MV, 5 μ m, 300 X 4.6 mm reversed phase column on 0.1% (wt/v) TFA with acetonitrile as eluent and a flow rate of 1.0 ml/min, gradient elution 1.25% acetonitrile/min. Preparatory reverse phase HPLC was performed on a Beckman HPLC with a Waters DeltaPak 25 X 100 mm, 100-mm C18 column equipped with a guard, 0.1% (wt/v) TFA, 0.25% acetonitrile/min. Water was obtained from a Millipore MilliQ water purification system, and all buffers were 0.2 μ m filtered. Reagent-grade chemicals were used unless otherwise stated. Oligonucleotides were synthesized by the Caltech Biopolymer Synthesis and Analysis Resource Center. Restriction enzymes were purchased from either New England Biolabs or Boeringer-Mannheim and used according to the manufacturer's protocols. [α - 32 P]-Thymidine-5'-triphosphate (≥ 3000 Ci. mmol) and [α - 32 P]-deoxyadenosine-5'-triphosphate (≥ 6000 Ci. mmol) were purchased from DuPont/NEN. [γ - 32 P]-Adenosine-5'-triphosphate (≥ 7000 Ci. mmol) was obtained from ICN.

Synthesis of polyamide- bis(dichloroethylamino)benzene conjugates

ImPy- β -ImPy-(R) H_2N γ -ImPy- β -ImPy- β -Dp (6). ImPy- β -ImPy-(R) H_2N γ -ImPy- β -ImPy- β -Pam-resin was synthesized in a stepwise fashion by Boc-chemistry manual solid phase protocols (21, 22). A sample of the resin was treated with neat (dimethylamino)-propylamine (2 ml), heated (55° C, 48 hours) and purified by reversed phase HPLC. ImPy- β -ImPy-(R) H_2N γ -ImPy- β -ImPy- β -Dp was recovered as a white powder upon

lyophilization of the appropriate fraction (18 mg, 10.4% recovery). UV (H_2O) λ_{max} (ϵ), 306 (69,500); ^1H NMR ($\text{DMSO}-d_6$): δ 10.97 (s, 1 H), 10.39 (s, 1 H), 10.25 (s, 1H), 10.00 (s, 1H), 9.97 (s, 1H), 9.90 (s, 1H), 9.24 (bs, 1 H), 8.15 (m, 5 H), 8.03 (bd, 3 H, $J = 6.1$ Hz), 7.47 (s, 1H), 7.42 (s, 1 H), 7.40 (s, 1 H), 7.35 (s, 1 H), 7.18 (m, 4 H), 7.01 (s, 1 H), 6.97 (s, 1H), 6.92 (s, 1 H), 6.90 (m, 2 H), 3.91 (m, 13 H), 3.75 (s, 12 H), 3.33 (m, 6 H), 3.24 (m, 2 H), 3.06 (q, 2 H, $J = 6.0$), 2.95 (quintet, 2 H, $J = 5.0$), 2.54 (t, 4 H, $J = 6.8$ Hz), 2.30 (t, 2 H, $J = 7.0$), 1.94 (m, 2 H), 1.68 (quintet, 2 H, $J = 7.3$ Hz); MALDI-TOF-MS (monoisotopic) $[\text{M} + \text{H}]$ 1381.6 (1381.6 calcd for $\text{C}_{62}\text{H}_{81}\text{N}_{26}\text{O}_{12}$).

ImPy- β -ImPy-(R)^{CHL} γ -ImPy- β -ImPy- β -Dp (2). To a solution of chlorambucil (5.7 mg, Fluka) in 20 μl DMF was added DCC (7.7 mg) and HOBT (2.7 mg). The solution was stirred for 1 hour. The OBt ester solution was added to ImPy- β -ImPy-(R)^{H₂N} γ -ImPy- β -ImPy- β -Dp (4 mg) in 400 μl DMF followed by 200 μl DIEA. The reaction mixture was stirred for 2 hours. TFA (150 μl) was added to the reaction mixture and the mixture was purified by reversed phase HPLC. ImPy- β -ImPy-(R)^{CHL} γ -ImPy- β -ImPy- β -Dp was recovered as a white powder upon lyophilization of the appropriate fraction (1 mg, 21% recovery). UV (H_2O) λ_{max} (ϵ) 306 (69,500); ^1H NMR ($\text{DMSO}-d_6$): δ 10.43 (s, 1H), 10.31 (s, 1 H), 10.30 (s, 1 H), 10.25 (s, 1 H), 10.09 (s, 1 H), 9.96 (s, 1 H), 9.52 (s, 1 H), 8.09, (m, 8 H), 7.47 (s, 2 H), 7.45 (s, 1 H), 7.40 (s, 1 H), 7.25 (m, 4 H), 7.05 (s, 2 H), 7.02 (s, 1 H), 6.96 (m, 4 H), 6.67 (s, 1 H), 6.64 (s, 1 H), 3.98 (s, 3 H), 3.95 (s, 6 H), 3.81 (s, 6 H), 3.79 (s, 1 H), 3.69 (s, 6 H), 3.52 (M, 8 H), 3.12 (q, 2 H, $J = 5.8$ Hz), 3.01 (quintet, 2 H, $J = 6.0$ Hz), 2.75 (d, 6 H, $J = 4.7$ Hz), 2.61 (m, 4 H), 2.37 t, (t, $J = 6.9$ Hz), 2.03 (t, 6 H, $J = 6.9$ Hz), 2.01 (m, 8 H), 1.74 (m, 4 H), 1.63 (m, 2 H), 1.47 (m, 2 H); MALDI-TOF-MS (monoisotopic) $[\text{M} + \text{H}]$ 1666.7 (1666.7 calcd for $\text{C}_{76}\text{H}_{98}\text{Cl}_2\text{N}_{27}\text{O}_{13}$).

ImPy- β -ImPy-(R)^{CHL(OH)²} γ -ImPy- β -ImPy- β -Dp (4). ImPy- β -ImPy-(R)^{CHL} γ -ImPy- β -ImPy- β -Dp (50 nmol) was dissolved in 0.1 NaOH (100 μ l). After 4 hours, hydrolysis was determined complete by analytical HPLC. The solution was neutralized with ammonium acetate and concentrated *in vacuo*.

Dp-CHL (5). To a solution of chlorambucil (304 mg) in 500 μ l DMF was added DCC (208 mg) and HOBt (143 mg). The solution was stirred for 1 hour. The OBt ester solution was added to (dimethylamino)-propylamine (63 μ l) in 500 ml DMF followed by 500 μ l DIEA. The reaction mixture was stirred for 2 hours. TFA (0.01% (wt/v), 6 ml) was added to the reaction mixture and the mixture was purified by reversed phase HPLC. Dp-CHL is recovered as a yellow oil upon lyophilization of the appropriate fraction (50 mg, 12.9% recovery). MALDI-TOF-MS (monoisotopic) [M + H] 388.4 (388.4 calcd for C₁₉H₃₁Cl₂N₃O).

Preparation of 3'- or 5'-³²P-labeled DNA restriction fragment

Plasmid pHIV-LTR (23) was digested with *EcoRI*, labeled at either the 3' or 5' end and digested with *HindIII*. The 241 base pair restriction fragment was isolated by nondenaturing gel electrophoresis. Chemical sequencing reactions were performed as described (28, 29). Standard techniques were employed for DNA manipulation (30).

Detection of heat-induced single-strand DNA cleavage

All reactions were executed in a total volume of 40 μ l. A polyamide stock solution or H₂O was added to an assay buffer containing radiolabeled restriction fragment (10,000 cpm), affording final solution conditions of 10 mM Tris•HCl, 10 mM KCl, 10 mM MgCl₂, 5 mM CaCl₂, pH 7.0, and either 1 pM-10 nM nitrogen mustard-polyamide

conjugate or no conjugate (for reference lanes). The solutions were equilibrated at 37° C for the desired time period. The reactions were stopped by the addition of 60 ml of a solution containing 0.6 M NaOAc, 12.5 mM EDTA, 0.15 mM (bp) calf thymus DNA, 2.0 M NaCl, 0.8 mg/ml glycogen, and ethanol precipitated. Reactions were resuspended in 40 ml 10 mM sodium citrate buffer, pH 7.2, and heated for 30 minutes at 95° C. The reactions were ethanol precipitated and resuspended in 1X TBE/80% formamide loading buffer, denatured by heating at 85° C for 10 min, and placed on ice. The reaction products were separated by electrophoresis on an 8% polyacrylamide gel (5% cross-link, 7 M urea) in 1X TBE at 2000 V. Gels were dried and exposed to a storage phosphor screen (Molecular Dynamics)(31).

Quantitative DNase I footprint titration experiments

All reactions were executed in a total volume of 400 µl. A polyamide stock solution or H₂O (for reference lanes) was added to an assay buffer containing radiolabeled restriction fragment (20,000 cpm) affording final solution conditions of 10 mM Tris•HCl, 10 mM KCl, 10 mM MgCl₂, 5 mM CaCl₂, pH 7.0, and either 1 pM-10 nM nitrogen mustard-polyamide conjugate or no conjugate (for reference lanes). The solutions were allowed to equilibrate at 22° C for 18 hours. Footprinting reactions were initiated by the addition of 4 µl of a DNase I stock solution (at the appropriate concentration to give about 55% intact DNA) containing 1 mM DTT and allowed to proceed for 7 min at 22° C. The reactions were stopped by the addition of 50 µl of a solution containing 1.25 M NaCl, 100 mM EDTA, 0.2 mg/ml glycogen, and 28 mM (bp) calf thymus DNA, and ethanol precipitated. Reactions were resuspended in 1 X TBE/80% formamide loading buffer,

denatured by heating at 85° C for 10 min, and placed on ice. The reaction products were separated by electrophoresis on an 8% polyacrylamide gel (5% cross link, 7M urea) in 1 X TBE at 2000 V for 1.5 h. Gels were dried and exposed to a storage phosphor screen (Molecular Dynamics).

Quantitation and data analysis

Data from the footprint titration gels were obtained using a Molecular Dynamics 400S PhosphorImager followed by quantitation using ImageQuant software (Molecular Dynamics). Background-corrected volume integration of rectangles encompassing the footprint sites and a reference site at which DNase I reactivity was invariant across the titration generated values for the site intensities (I_{site}) and the reference intensity (I_{ref}). The apparent fractional occupancy (θ_{app}) of the sites were calculated using the following equation:

$$\theta_{\text{app}} = 1 - \frac{I_{\text{site}}/I_{\text{ref}}}{I_{\text{site}}^{\circ}/I_{\text{ref}}^{\circ}} \quad (1)$$

where I_{site}° and I_{ref}° are the site and reference intensities, respectively, from a control lane to which no polyamide was added. The ($[L]_{\text{tot}}$, θ_{app}) data points were fit to a Langmuir binding isotherm (eq 2, $n=1$) by minimizing the difference between θ_{app} and θ_{fit} , using the modified Hill equation:

$$\theta_{\text{fit}} = \theta_{\text{min}} + (\theta_{\text{max}} - \theta_{\text{min}}) \frac{K_a^n [L]_{\text{tot}}^n}{1 + K_a^n [L]_{\text{tot}}^n} \quad (2)$$

where $[L]_{\text{tot}}$ is the total polyamide concentration, K_a is the equilibrium association constant, and θ_{min} and θ_{max} are the experimentally determined site saturation values when

the site is unoccupied or saturated, respectively. The data were fit using a nonlinear least-squares fitting procedure with K_a , θ_{\max} , and θ_{\min} as the adjustable parameters. All acceptable fits had a correlation coefficient of $R > 0.97$. Five sets of data were used in determining each association constant. All lanes from each gel were used unless visual inspection revealed a data point to be obviously flawed relative to neighboring points.

Acknowledgments

We are grateful to the National Institutes of Health General Medical Sciences for research support, and to the Ralph M. Parsons Foundation for a predoctoral fellowship to N. R. W.

References

1. Trauger, J. W., Baird, E. E., and Dervan, P. B. (1996) *Nature* 382, 559-561.
2. Swalley, S. E., Baird, E. E., and Dervan, P. B. (1997) *J. Am. Chem. Soc.* 119, 6953-6961.
3. Turner, J. M., Baird, E. E., and Dervan, P. B. (1997) *J. Am. Chem. Soc.* 119, 7636-7644.
4. Trauger, J. W., Baird, E. E., and Dervan, P. B. (1998) *Angewandte Chemie-International Edition* 37, 1421-1423.
5. Turner, J. M., Swalley, S. E., Baird, E. E., and Dervan, P. B. (1998) *J. Am. Chem. Soc.* 120, 6219-6226.
6. Wade, W. S., Mrksich, M., and Dervan, P. B. (1992) *J. Am. Chem. Soc.* 114, 8783-8794.
7. Mrksich, M., Wade, W. S., Dwyer, T. J., Geierstanger, B. H., Wemmer, D. E., and Dervan, P. B. (1992) *Proc. Natl. Acad. Sci. U. S. A.* 89, 7586-7590.
8. White, S., Baird, E. E., and Dervan, P. B. (1997) *Chem. Biol.* 4, 569-578.
9. Kielkopf, C. L., Baird, E. E., Dervan, P. D., and Rees, D. C. (1998) *Nat. Struct. Biol.* 5, 104-109.
10. Kielkopf, C. L., White, S., Szewczyk, J. W., Turner, J. M., Baird, E. E., Dervan, P. B., and Rees, D. C. (1998) *Science* 282, 111-115.
11. White, S., Szewczyk, J. W., Turner, J. M., Baird, E. E., and Dervan, P. B. (1998) *Nature* 391, 468-471.
12. Dervan, P. B., and Burli, R. W. (1999) *Current Opinion in Chemical Biology* 3, 688-693.

13. Trauger, J. W., Baird, E. E., Mrksich, M., and Dervan, P. B. (1996) *J. Am. Chem. Soc.* **118**, 6160-6166.
14. Swalley, S. E., Baird, E. E., and Dervan, P. B. (1997) *Chem.-Eur. J.* **3**, 1600-1607.
15. Trauger, J. W., Baird, E. E., and Dervan, P. B. (1998) *J. Am. Chem. Soc.* **120**, 3534-3535.
16. Gottesfeld, J. M., Neely, L., Trauger, J. W., Baird, E. E., and Dervan, P. B. (1997) *Nature* **387**, 202-205.
17. Dickinson, L. A., Gulizia, R. J., Trauger, J. W., Baird, E. E., Mosier, D. E., Gottesfeld, J. M., and Dervan, P. B. (1998) *Proc. Natl. Acad. Sci. U. S. A.* **95**, 12890-12895.
18. Mattes, W. B., Hartley, J. A., and Kohn, K. W. (1986) *Nucleic Acids Research* **14**, 2971-2987.
19. Zhi-Fu, T., Fujiwara, T., Saito, I., and Sugiyama, H. (1999) *J. Am. Chem. Soc.* **121**, 4961-4967.
20. Baraldi, P. G., Cacciari, B., Guiotto, A., Romagnoli, R., Zaid, A. N., and Spalluto, G. (1998) *Current Pharmaceutical Design* **4**, 249-276.
21. Herman, D. M., Baird, E. E., and Dervan, P. B. (1998) *J. Am. Chem. Soc.* **120**, 1382-1391.
22. Baird, E. E., and Dervan, P. B. (1996) *J. Am. Chem. Soc.* **118**, 6141-6146.
23. Jones, K. A., and Peterlin, B. M. (1994) *Annual Review of Biochemistry* **63**, 717-743.

24. Reynolds, V. L., Molineux, I. J., Kaplan, D. J., Swenson, D. H., and Hurley, L. H. (1985) *Biochemistry* 24, 6228-6237.
25. Wyatt, M. D., Lee, M., Garbiras, B. J., Souhami, R. L., and Hartley, J. A. (1995) *Biochemistry* 34, 13034-13041.
26. Brooks, N., Hartley, J. A., Simpson, J. E., Wright, S. R., Woo, S., Centioni, S., Fontaine, M. D., McIntyre, T. E., and Lee, M. (1997) *Bioorganic & Medicinal Chemistry* 5, 1497-1507.
27. Kundu, G. C., Schullek, J. R., and Wilson, I. B. (1994) *Pharmacology Biochemistry and Behavior* 49, 621-624.
28. Maxam, A. M., and Gilbert, W. (1986) *Molecular Biology* 20, 461-509.
29. Iverson, B. L., and Dervan, P. B. (1987) *Nucleic Acids Research* 15, 7823-7830.
30. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
31. Johnston, R. F., Pickett, S. C., and Barker, D. L. (1990) *Electrophoresis* 11, 355-360.

CHAPTER 7

Alkylation, Toxicity and Cell Permeability of Chlorambucil-Polyamide Conjugates

Abstract

The combination of DNA binding pyrrole/imidazole polyamides with nonspecific alkylator, chlorambucil, creates a class of sequence-specific DNA alkylating compounds. Since the compounds exhibit good specificity and yield of DNA alkylation, we investigated their activity in biological systems. A chlorambucil-polyamide conjugate displays enhanced *in vitro* cross-linking, alkylation and inhibition of replication of SV40 DNA when compared to unmodified chlorambucil. Additionally, the compound is shown to alkylate its target site in a HIV promoter DNA sequence that has been integrated into genomic DNA in cells and in cell free conditions. A dye labeled derivative showed increased nuclear localization in three cell types relative to a dye-labeled polyamide with the alkylator moiety when examined by confocal microscopy. Most interestingly, the conjugate showed markedly different toxicity for several cell lines. The evidence of nuclear localization further supports the proposed mechanism of gene regulation of polyamides and provides a lead for derivatization of polyamides to increase nuclear localization.

Introduction

A general method for specifically and efficiently alkylating and cutting predetermined DNA sequences with cell permeable molecules could be a useful tool for regulation of gene expression. An initial step toward this goal has been made by the creation of a class of molecules combining pyrrole/imidazole polyamides which bind predetermined DNA sequences with high affinity and specificity with the well-characterized DNA alkylator, chlorambucil (CHL) (1). Bis (dichloroethylamino) benzene derivatives were attached to the γ -turn of an eight-ring hairpin polyamide targeted to the HIV promoter and found to bind and selectively alkylate the desired target sites in the HIV promoter at subnanomolar concentrations. Since the chlorambucil-polyamide conjugates demonstrated that DNA damage could be targeted to specific DNA sequences by polyamides, we were interested in further investigating the DNA interactions and biological activities of this class of compounds.

We describe here several studies of the biological activity of polyamide-chlorambucil conjugate **1**. First data will be presented on *in vitro* cross-linking, alkylation and replication of SV40 DNA by compound **1** and unmodified chlorambucil. Next, *in vitro* vs. *in vivo* alkylation by compound **1** of a stably integrated HIV promoter will be compared. The toxicity of the compound will also be investigated in several cell lines. Finally, cell trafficking of a BODIPY-dye-polyamide conjugate studied by confocal microscopy will be examined.

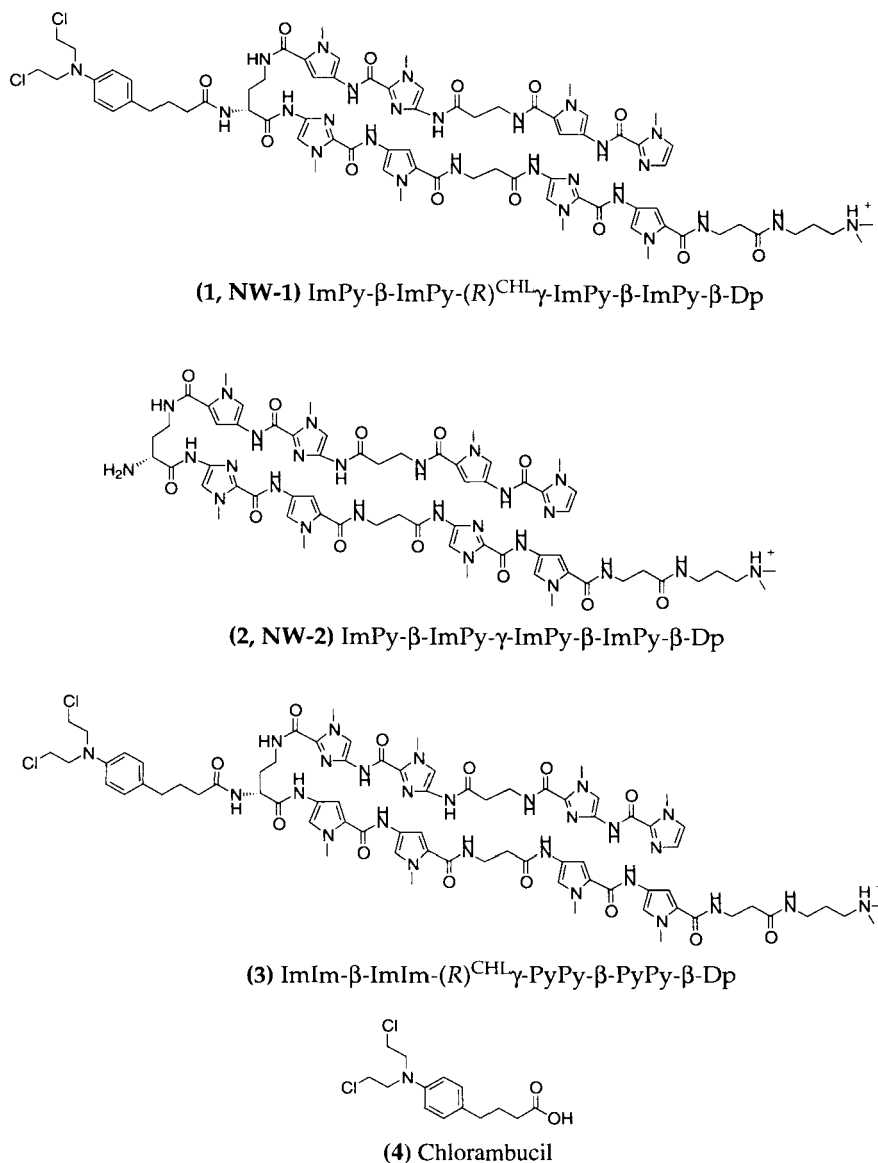


Figure 7.1. Structures of polyamides 1-3 and chlorambucil 4.

Results and Discussion

Induction of DNA interstrand cross-links *in vitro*

Methods used for the detection of DNA interstrand cross-links often involve the use of harsh procedures (high temperature or alkali) to separate single strands in linear plasmid DNA. As the stability of the chlorambucil-induced DNA alkylations/cross-links could be compromised under such conditions, we employed a “mild” methods based on

heating the linear DNA in 65% formamide to 55° C for 10 min. under these conditions the control DNA denatured and unstable cross-links were reasonably conserved (2).

The dependence of **NW-1**-induced DNA cross-linking on incubation time was investigated by treating linear SV40 DNA with 10 μ M drug for up to 4 h before ethanol precipitation and gel analysis of cross-link formation. As shown in Figure 7.2, a level of cross-links increased rapidly up to ~1 h, and after which began to plateau and the damage to DNA (low molecular “smear”) started to prevail (data not shown). Thus, 1 h incubation with **NW-1** was assumed to be optimal at this concentration for cross-linking studies.

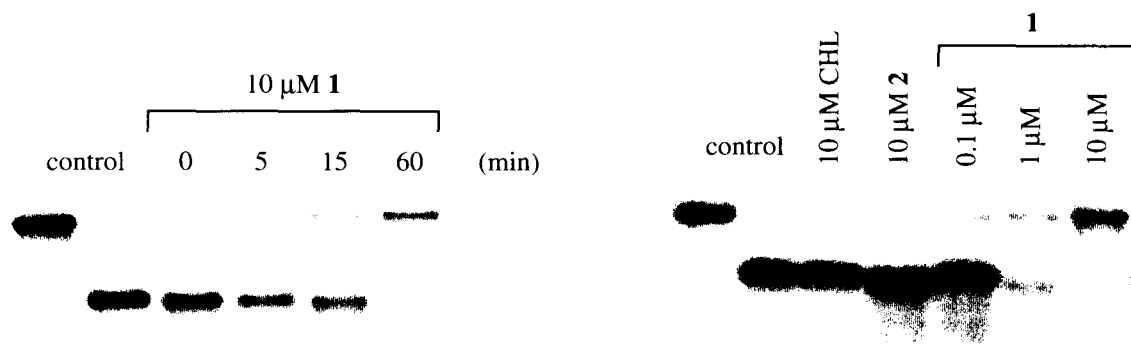


Figure 7.2. (A) Time dependence of DNA interstrand cross-link induction by chlorambucil polyamide conjugate **1**. SV40 DNA was incubated with 10 μ M **1**, purified and resolved on agarose gels. (B) Concentration dependence of cross-link induction in 1 hour, with CHL and polyamide **2** as controls.

Once the treatment time had been selected, the concentration dependence of interstrand cross-linking was studied. Linear SV40 DNA was treated for 1 h with 0.1, 1 and 10 μ M **NW-1**, then purified and analyzed on agarose gel. As shown in Figure XB, DNA treated with drug showed a concentration-dependent increase in DNA migration. At 10 μ M, ~75% of total DNA was cross-linked. The ability of **NW-1** and its parent compounds (**NW-2** and **CHL**) to induce DNA interstrand cross-links and/or alkylation in isolated DNA was compared using the conditions established above (1 h treatment with

up to 10 μ M drug). The results clearly show that **NW-1** is more efficient than freely-diffusing chlorambucil. Additionally, the polyamide compound lacking the alkylator was unable to preserve double-strand DNA under the employed “mild” denaturing conditions, thus excluding possibility of creating noncovalent “virtual” cross-links. It should be noted for this experiment and ensuing experiments that the half-life of DNA alkylation by polyamide **1** is ~ 5 h (*1*). Although the short incubation times (1 h) are sometimes necessary due to instability of the systems, the results of these studies are not a true reflection of the potency of the compound. Efficient alkylation and cross-linking likely occurs at lower concentrations and higher yields at 12 h equilibration times.

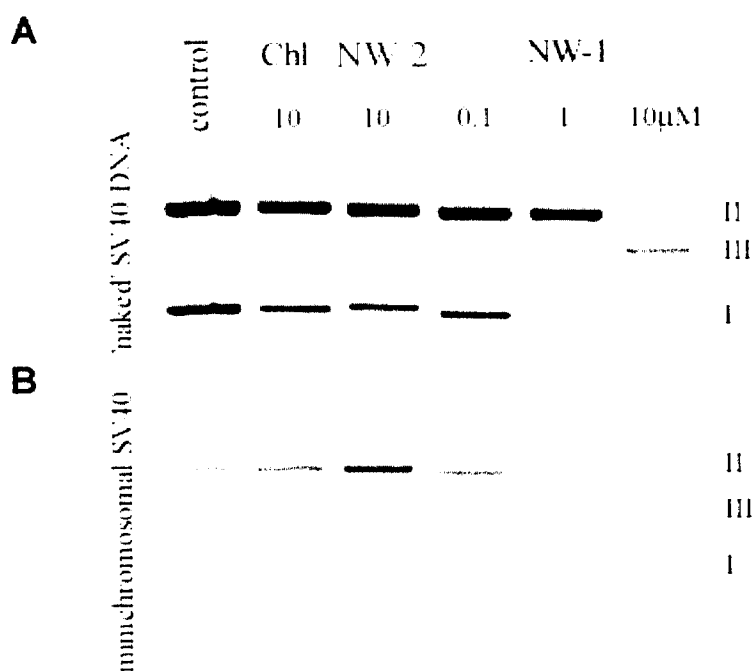


Figure 7.3. Comparison of in vitro SV40 DNA and SV40 minichromosomes alkylation and cross-linking by polyamides **1** and **2** and Chl after 1 h drug treatment. (A) Alkylation of SV40 DNA after 1 h treatment with 0.1, 1 and 10 μ M of Chl, 1 and 2. (B) Alkylation of SV40 minichromosomal DNA after 1 h treatment with polyamides **1** and **2** and Chl. SV40 DNA was drug-treated, purified, then resuspended in TE and heated at 70° C for 2 h to convert DNA alkylation lesions to strand breaks.

The drug-induced DNA alkylations were converted to DNA strands by heating at 70° C for 2 h (Figure 7.3). Both plasmid SV40 DNA and minichromosomal SV40 DNA were treated with the studied compounds (3-5), allowing a comparison between the drug's action upon "naked" and protein protected DNA. The damage observed of plasmid and minichromosomal DNA was similar. NW-1 could efficiently alkylate DNA at concentration $\leq 10 \mu\text{M}$, whereas the not cleavage was observed with CHL.

Inhibition of SV40 DNA replication in cell-free *in vitro* system

SV40 DNA replication was used to examine the functional effects of NW-1-induced DNA modifications (6). DNA was incubated with the compound at 3 and 0.3 μM concentrations, and then extensively purified, completely losing its replication template properties. SV40 DNA treated with as low as 0.3 μM NW-1 showed ~25% inhibition of replication activity.

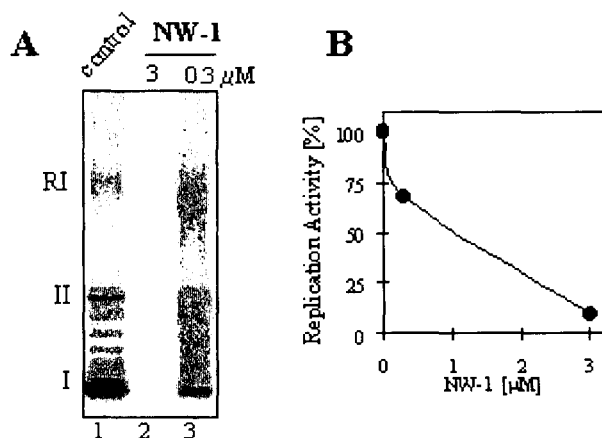


Figure 7.4. Polyamide 1 induced inhibition of SV40 DNA replication in cell-free system using drug-treated DNA template. Reaction mixtures containing 30 μg of cell extracts, 50 ng of SV40 DNA template and 0.6 μg of SV40 T-antigen were incubated at 37° C for 1 h. Replication products were isolated, resolved on 0.8% neutral agarose gels (A), visualized by autoradiography, analyzed and quantitated (B). Positions of plasmid form I, II and replication intermediates (RI) are indicated.

***In Vitro* vs. *In Vivo* Alkylation of Genomic DNA**

One criticism of the observed inhibition of HIV-1 transcription by hairpin polyamides (7) is that inhibition in cells was inferred from viral replication rather than from direct transcript levels. Potentially polyamides affect replication by downregulation of other genes or interruption of another cellular process. Alkylation by a chlorambucil-polyamide conjugate provides a direct chemical readout that demonstrates that the ligand can traffic through the cell membrane, into the nucleus, and binds the desired target sequence. CEM cells with a stably integrated HIV-1 promoter attached to a GFP reporter gene were treated with 500 nM of **1** or control **3** for 24 h. Following isolation of the genomic DNA and heat induced cleavage by piperidine, cleaved DNA in the HIV promoter region was amplified by ligation-mediated PCR with a P-32 labeled primer. For comparison, genomic DNA was treated with either of the compounds and worked up identically. No alkylation was observed in either case when cells or DNA were treated for only 1 h and worked up. The fragments were separated on a sequencing gel for analysis.

Alkylation was observed at match sites adjacent to the TATA Box and ETS sites, as well as at a mismatch site (see Figure 7.5). This compares favorably to alkylation sites observed on a restriction fragment that contained the same HIV promoter DNA sequence. Alkylation yields were much lower, though, since significant cleavage was observed at low nanomolar concentrations under optimized *in vitro* conditions on a 241 bp restriction fragment (1). Higher concentrations and lower alkylation efficiency are expected, though, since many other match sites are present in the genomic DNA. More importantly, the roughly equal alkylation of the match site, whether the genomic DNA is in the nucleus

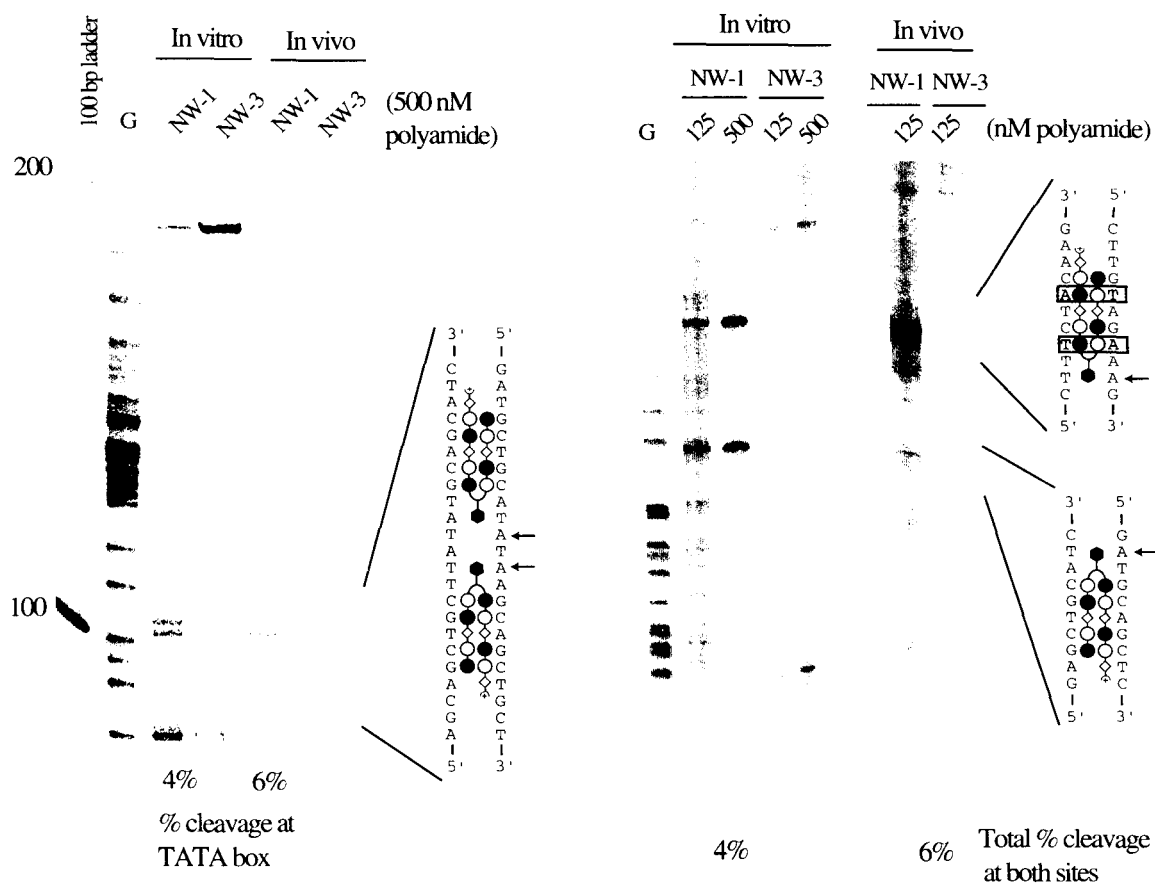


Figure 7.5. Comparison of in vitro and in vivo alkylation by chlorambucil-polyamide conjugate 1. For in vivo results, CEM cells were treated with 500 nM of 1 overnight. The genomic DNA was isolated, treated with piperidine and heated to induced DNA strand cleavage. The cleaved DNA was amplified by ligation-mediated PCR. Storage phosphor autoradiograms: (A) Coding strand for the TATA box. (B) Non-coding strand for the Ets-1 site. Alkylation sites are shown to the right of the gels.

or in solution, potentially shows that the polyamide can traffic into the nucleus and bind it's designed target site in the nucleus.

A Survey of Cellular Growth Inhibition

Since researchers in the Baltimore and Beerman groups had observed different toxicities used polyamide-chlorambucil conjugates (toxic at 10 nM in 293T cells and not toxic in BSC-1 cells, respectively), it seemed prudent to look at a panel of cell lines to scan for biological activity. Interestingly, NW-1 has vastly different toxicity profiles for

different cell lines ranging from inhibition of cell growth at low nanomolar concentrations to no observed effects of growth at concentrations of 2 μM . The compound demonstrated highest toxicity to T-cells, which correlates well with success at inhibition of transcription by hairpin polyamides in T-cells. Like most alkylating agents NW-1 was observed to block cells in G2 phase, resulting in the observed inhibition of cell growth.

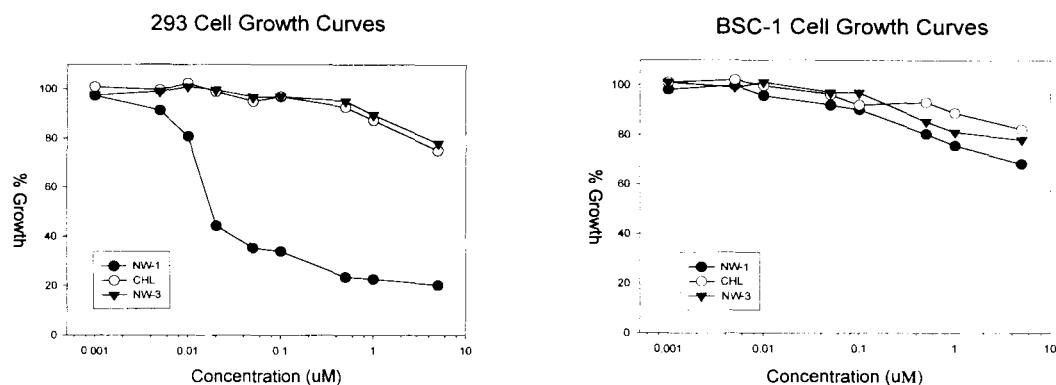


Figure 7.6. Inhibition of cell growth after treatment for 4 days with compounds **1** and **2** and Chl.

Table 7.1. Cell lines sensitive to NW-1 ($\text{IC}_{50} < 1000 \text{ nM}$)

Cell Line	Cell line description	IC_{50} (nM)
Jurkat	Human R cell leukemia (suspension)	~2.2
293	Human transformed embryonic kidney cells	~15
H460	Human non-small-cell lung carcinoma (NSCLC)	~18.6
Hela S3	Human epitheloid carcinoma	~200
A2780-DX5	Ovarian cancer cells - resistant	~737

Table 7.2. Cell lines NOT sensitive to NW-1 ($\text{IC}_{50} > 2000 \text{ nM}$)

Cell line	Cell line description
BSC-1	African green monkey kidney cells
HCT-116	Human colon carcinoma
LCC6-MDR	Breast Cancer cells - resistant
DLD-1	Human colon carcinoma
A549	Human non-small-cell lung carcinoma (NSCLC)

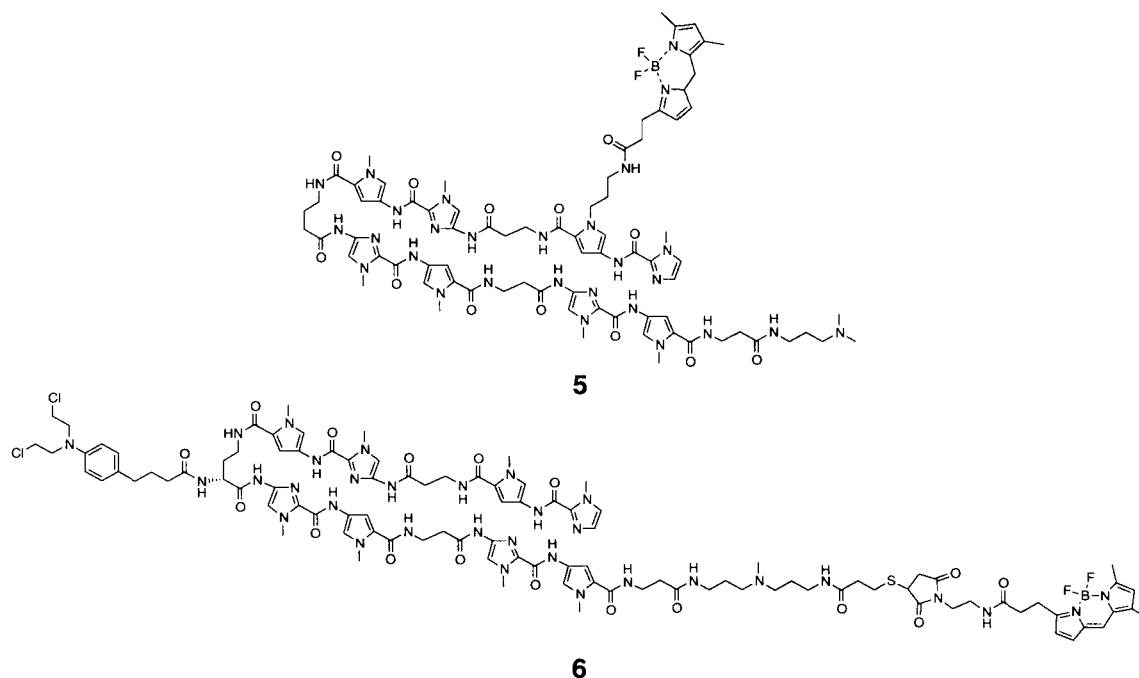


Figure 7.7. Structures of BODIPY labeled polyamides for cell permeability studies.

Cell Permeability by a Chlorambucil-BODIPY-polyamide conjugate

After initial successes of gene downregulating in cells with hairpin polyamides, it was observed that in several cases polyamides were successful in inhibiting transcription of a specific gene *in vitro* but was unsuccessful when tested in cells. While other explanations are possible, such as lack of specificity in a genomic DNA context, we speculated that lack of inhibition in a cellular system may be due to polyamides not trafficking to the nucleus. In order to investigate this hypothesis, a series of polyamide dye conjugates was synthesized and their cell trafficking properties were investigated by confocal microscopy. When polyamide-dye conjugates were added to the cell media, the ligands were observed in the nucleus in only a few cell types. When comparing the toxicity of chlorambucil-polyamide conjugates to the cell permeability data with polyamide-dye conjugates, it is striking that cell lines, such as 293T, alkylator conjugates

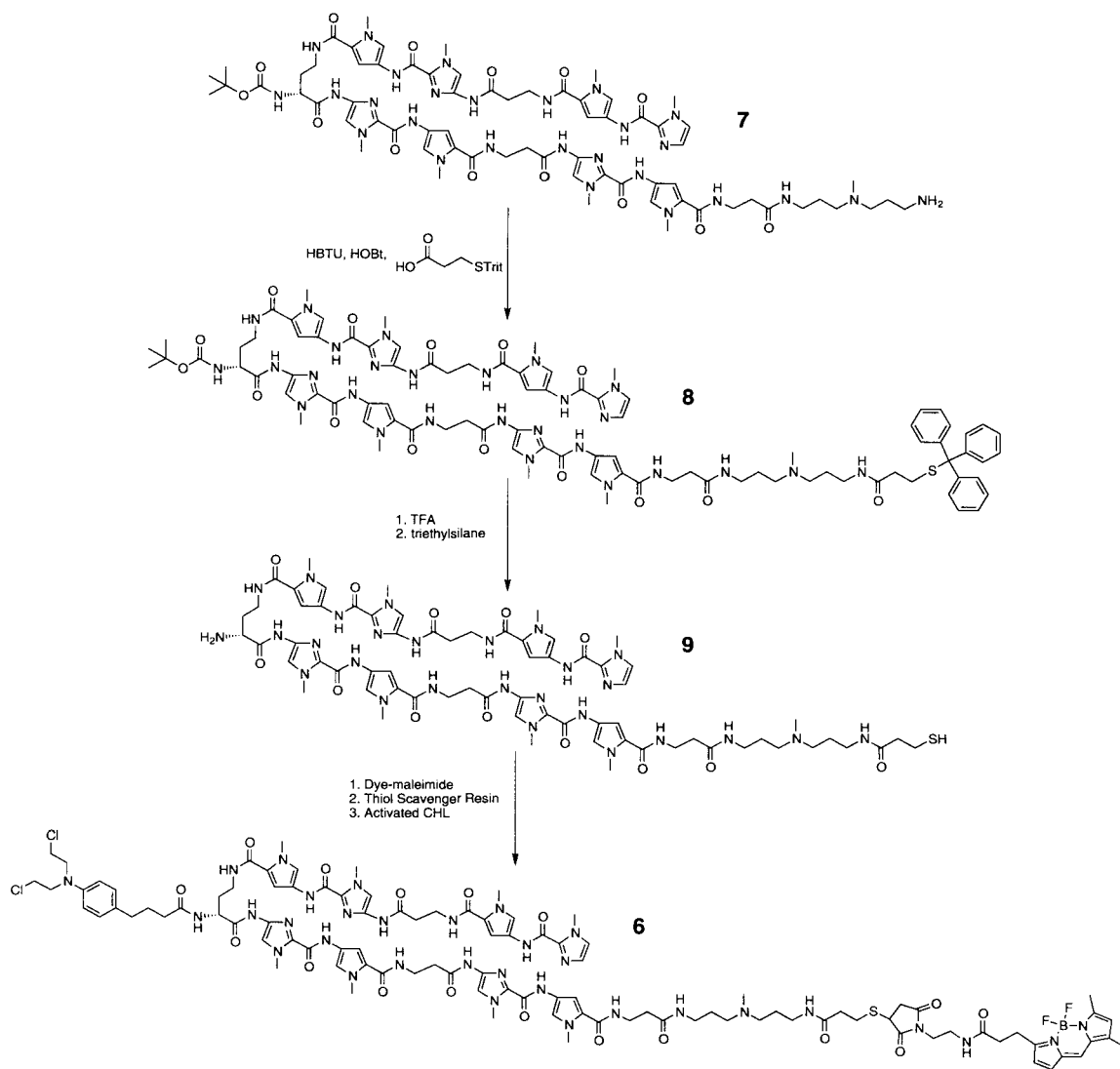


Figure 7.8. Synthesis of dye-alkylator-polyamide conjugate **6**.

are extremely toxic, yet a dye conjugate of the same parent polyamide is observed not to localize in the nucleus. In light of the conflicting data a chlorambucil-BODIPY-polyamide conjugate was synthesized to investigate whether the chlorambucil moiety would alter cell trafficking properties.

After synthesis of the polyamide by previously reported solid phase synthesis methods with Boc-protected monomers (**8**), the Fmoc protecting group on the DABA amine was deprotected with piperidine and replaced with a Boc group with Boc

anhydride. Cleavage from resin with the triamine and purification by prep HPLC provide compound **7** for conjugation. Coupling of S-trityl- β -propionic acid followed by deprotection of the trityl group and purification by Sep-Pak produced compound **9** suitably functionalized for a one-pot addition of BODIPY and chlorambucil. A slight excess of BODIPY-maleimide was addition to **9** with 0.5 eq of DIEA. Excess maleimide was removed from the reaction mixture with a thiol scavenger resin followed by filtering off the resin. An excess DCC, HOBt activated chlorambucil was added and the resulting product was purified by preparatory HPLC. Identity and purity of all compounds were determined by HPLC and MALDI-TOF mass spectrometry (Figure 7.8).

Several cell lines were treated with 5 μ M of compounds **5** and **6** and incubated overnight. Cell death of the dye-chlorambucil-polyamide conjugate was not observed at these concentrations, likely because the conjugation of the dye decreases DNA binding affinity and therefore alkylating potency. Examination of the cells by confocal microscopy revealed that addition of chlorambucil did not improve nuclear localization except in a few cases (Figure 7.9). Most notably polyamide **6** was localized in the nucleus in NB4 cells, whereas compound **5** was only observed in the cytoplasm. Nuclear localization in CEM cells and primary T-cells was also significantly increased for compound **6**. Surprisingly, the alkylator was not observed in the nucleus in 293T cells, in which compound **1** showed low nanomolar toxicity. One can postulate that this discrepancy occurs because the dye alters nuclear localization or that toxicity is due a mechanism that does not involve DNA. Alternatively, the amount of alkylator in the nucleus necessary for inhibition of cell growth could be too low to be observed by confocal microscopy.

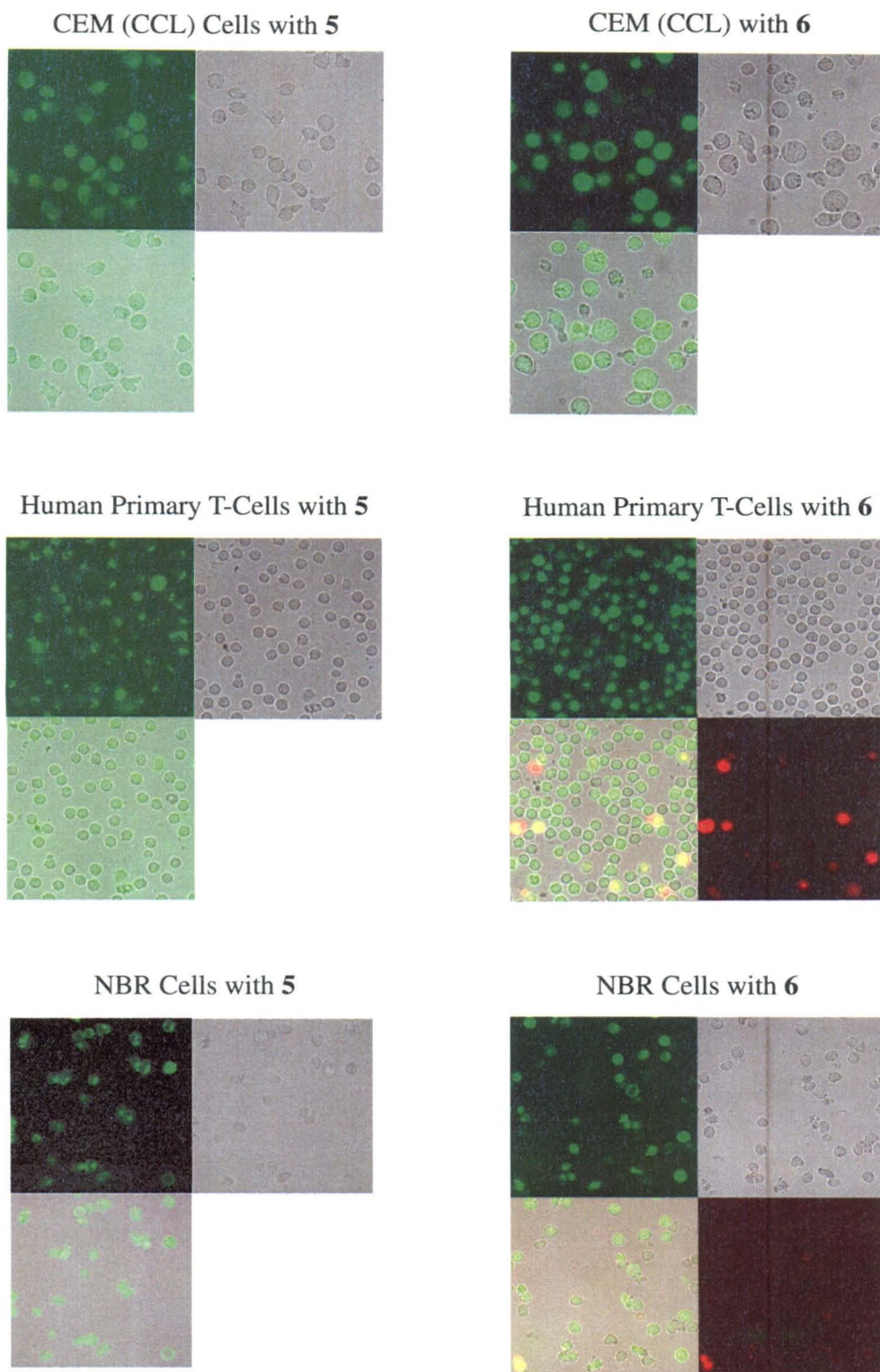


Figure 7.9. Imaging cell trafficking of polyamide dye conjugates by confocal microscopy. (upper left) Fluorescent image (upper right) bright field image (bottom left) merged images (bottom right) Dead cells stained with Sytox Orange.

Conclusion

The variable toxicity to cells is likely the results of several factors. The differential cell uptake by polyamide-dye conjugates points to the obvious possibility that the polyamide-alkylators have different cell trafficking properties for each of the cell lines. With the current data, though, it must be considered that some cell lines may be more sensitive to DNA damage. Studies in recent years have demonstrated that cell lines have different thresholds of apoptosis. An additional consideration is that certain cell lines are able to more efficiently repair DNA damage induced by the alkylators. The sequence selectivity of the polyamide may target DNA sequences that are more vital to the sensitive cell lines. While there is no definitive answer at this time to explain the efficacy of the alkylator, the observed pharmacological properties of this new class of compounds represent a beginning of line of scientific inquiry. As polyamides with higher specificity for DNA within a genomic context are developed, alkylator polyamide conjugates may provide a means for “genetic microsurgery,” whereby undesired gene segments or integrated viral DNAs could be removed from a host genome *in vivo*. At a minimum these results provide a clue that may lead future investigators to modify polyamides in such a way to improve cell-tracking properties, perhaps even grant cell-type specific nuclear localization.

Acknowledgments. Thank you to Jason Belitsky and Bobby Arora for the confocal microscopy. Liliane Dickinson and Joel Gottesfeld performed the investigation of DNA alkylation of genomic DNA, while Michael Wang and Terry Beerman conducted the experiments with SV40 and inhibition of cell growth.

Experimental

ImPy- β -ImPy-(*R*)^{NHBOC} γ -ImPy- β -ImPy- β -Trit (7) ImPy- β -ImPy-(*R*)^{NHFmoc} γ -ImPy- β -ImPy- β -PAM Resin was synthesized in a stepwise fashion manual solid phase methods from Boc- β -PAM resin (600 mg, 0.75 mmol/g). A portion of the resin (200 mg) was deprotected with 20% piperidine in DMF. The amine was reprotected with Boc₂O and DIEA. The resin was washed, dried, and treated with 3,3'-diamino-*N*-methyldipropylamine for 12 hours at 60° C. Purification as previously reported yielded ImPy- β -ImPy-(*R*)^{NHBOC} γ -ImPy- β -ImPy- β -Tr **7** (28 μ mol, 18.6% recovery). UV (H₂O) λ_{max} 310 (68,000) MS (MALDI-TOF): [M+H]⁺ 1524.8 (C₆₉H₇₆N₂₀O₁₁⁺, [M+H]⁺, calc. 1524.7)

ImPy- β -ImPy-(*R*)^{NHBOC} γ -ImPy- β -ImPy- β -Tr- β -SHTr (8) *S*-trityl- β -propionic acid (35 mg, 100 μ mol) and HBTU (38 mg, 100 μ mol) were dissolved in DMF (100 μ l) followed by addition of DIEA (50 μ l). After 15 minutes, 10 μ l of the activated acid was added to polyamide **7**, dissolved in DMF (200 μ l). After 1 hour and room temperature, the reaction was determined complete by analytical HLPC. The product was purified on a C8 Sep-Pak cartridge (2 g) with acetonitrile and 0.1% aqueous TFA. Pure fractions were identified by analytical HPLC and lyophilized to give a white solid (3.2 μ mol). UV (H₂O)

λ_{\max} 310 (68,000) MS (MALDI-TOF): $[M+H]^+$ 1854.9 ($C_{91}H_{112}N_{27}O_{14}^+$, $[M+H]^+$, calc. 1855.0).

ImPy- β -ImPy-(*R*)^{NH₂}- γ -ImPy- β -ImPy- β -Tr- β -SH (9) Polyamide **8** (3.2 μ mol) was dissolved in neat TFA (100 μ l). After 5 minutes triethyl silane was added to the bright yellow solution until the color dissipated (~20 μ l). TFA was removed *in vacuo*, the residue was dissolved in 20% acetonitrile/80% 0.1% TFA, and the product was purified on a C8 Sep-Pak cartridge (2 g) with acetonitrile and 0.1% aqueous TFA. Pure fractions were identified by analytical HPLC and lyophilized to give a white solid (2.4 μ mol). UV (H_2O) λ_{\max} 310 (68,000) MS (MALDI-TOF): $[M+H]^+$ 1512.7 ($C_{67}H_{90}N_{27}O_{13}^+$, $[M+H]^+$, calc. 1512.8).

Polyamide (6). BODIPY-maleimide (2.6 μ mol in 20 μ l of DMF) was added to a solution of polyamide **9** (2.4 μ mol) dissolved in DMF (200 μ l) followed by addition of DIEA (2 μ l). After the reaction was judged complete by analytical HPLC, thiol scavenger resin (20 mg) was added. In a separate tube, chlorambucil (46 mg), DCC (31 mg) and HOBT (22 mg) were dissolved in DMF. After 30 minute the resin was filtered off and 20 μ l of the activated chlorambucil solution was added to the polyamide-dye solution. The reaction mixture was mixed at room temperature for 1 hour and then preparatory reverse phase HPLC afforded polyamide **6** upon lyophilization of the appropriate fractions (900 nmol 34.6% recovery). UV (H_2O) λ_{\max} 310 (68,000) MS (MALDI-TOF): $[M+H]^+$ 2211.9 ($C_{101}H_{128}BCl_2F_2N_{32}O_{17}^+$, $[M+H]^+$).

References

1. Wurtz, N. R., and Dervan, P. B. (2000) *Chemistry & Biology* 7, 153-161.
2. Hartley, J. A., Berardini, M. D., and Souhami, R. L. (1991) *Analytical Biochemistry* 193, 131-134.
3. McHugh, M. M., Woynarowski, J. M., Mitchell, M. A., Gawron, L. S., Weiland, K. L., and Beerman, T. A. (1994) *Biochemistry* 33, 9158-9168.
4. Grimwade, J. E., and Beerman, T. A. (1986) *Molecular Pharmacology* 30, 358-363.
5. Woynarowski, J. M., McHugh, M. M., Gawron, L. S., and Beerman, T. A. (1995) *Biochemistry* 34, 13042-13050.
6. McHugh, M. M., Kuo, S. R., Walsh-O'Beirne, M. H., Liu, J. S., Melendy, T., and Beerman, T. A. (1999) *Biochemistry* 38, 11508-11515.
7. Dickinson, L. A., Gulizia, R. J., Trauger, J. W., Baird, E. E., Mosier, D. E., Gottesfeld, J. M., and Dervan, P. B. (1998) *Proceedings of the National Academy of Sciences of the United States of America* 95, 12890-12895.
8. Baird, E. E., and Dervan, P. B. (1996) *Journal of the American Chemical Society* 118, 6141-6146.